Powerful men on top:

Stereotypes interact with metaphors in social categorizations

Natalia Zarzeczna, Ulrich von Hecker, Travis Proulx, and Geoffrey Haddock

School of Psychology, Cardiff University

Acknowledgements

We would like to thank our colleagues from the Cardiff University Social Psychology lab for their feedback concerning this research. Also, we are grateful to research assistants, especially Anna Błaszczak and Paul Goesmann, for help in data collection. Finally, we would also like to thank reviewers for their useful input in an earlier version of this manuscript.

This manuscript is partially based on the PhD thesis of the first author.

The research was supported by an Economic and Social Research Council Doctoral Fellowship to the first author.

Our data and materials are publicly available at: https://doi.org/10.17605/OSF.IO/NEQH3

Abstract

We examined whether people can simultaneously apply two cognitive strategies in social categorizations. Specifically, we tested whether stereotypes concerning social power of gender categories interact with metaphoric power-space links. Based on the conceptual blending perspective suggesting that semantically consistent concepts acquire each other's properties, we predicted the following: Given that stereotypes create expectations linking gender with power, and metaphorically power is linked with vertical space, the conceptual blend of gender-power-space would invoke representations of male targets at the top vertical position when categorizing them as powerful, whilst female targets at the bottom when categorizing them as powerless. Across six studies, we show that the concept of gender is simulated spatially when people attribute power to male, but not female, targets. The predicted power-gender blending involved simulations of men judged as powerful when presented in upper location as opposed to women judged as powerful in upper location and men judged as powerful in lower location. Our hypothesis was further corroborated using pupillometry to assess pre-conscious processing, whereby stereotypically inconsistent orientations of gender and power evoked pupillary markers indicative of surprise. Our studies suggest that gender-power stereotypic expectations interact with the power-space metaphor in social categorizations.

Keywords:

Spatial processing; Power; Stereotypes; Pupillometry, Expectancy violation

Word Count: 15 742

Public Significance Statement:

These experiments demonstrate that people can simultaneously use more than one mental strategy to facilitate their decision-making about social groups. The application of stereotypes that serve as mental shortcuts in social categorizations is combined with other mental tools, such as linguistic metaphors. People combine stereotypic expectations of gender categories (men are powerful) with metaphoric associations (powerful are at the top) and as a result, men are mentally represented in an upper vertical location. Hence, people's responses are facilitated when categorizing a man, but not a woman, in an upper vertical location. To reduce stereotyping in social categorizations, it is essential to address the impact of supporting mechanisms that interact with stereotypes.

Powerful men on top:

Stereotypes interact with metaphors in social categorizations

"Space: Literally it means *nothing*, a vacuum between stars and planets, but by the same token it means *everything*. It's what connects all our worlds." - Captain Janeway, Star Trek Voyager[1]

While physical space exists in the distance between objects, space is also a medium that connects abstract ideas within our *mental* worlds. It provides context for our daily lives and interactions with other individuals. From our early days, we are exposed to physical spatial distances and orientations that are experienced through perceptual and motor capabilities. These experiences help shape our physical and mental development (Smith & Gasser, 2006). Further, abstract concepts can be illustrated or understood by concrete concepts (Lakoff & Johnson, 1999). This process can be seen in linguistic metaphors. For example, we often refer to powerful people as being "on the top of the pyramid" or the "top dog." This *control/power is up* metaphor indicates that it is literally easier to control/exert pressure or influence on someone when we are physically positioned at the top as opposed to the bottom. In the context of this framework, it is possible that metaphors do not simply reflect thought processes about linking abstract concepts with concrete ones but serve as a tool that guide the development of the links between concreteness and abstraction (Jamrozik, McQuire, Cardillo, & Chatterjee (2016). As such, space might not only connect our *physical* worlds; it also seems to connect abstract ideas in our *mental* worlds.

[1] The quote was obtained from: Quotes. (n.d.). Retrieved June 18, 2018, from https://www.imdb.com/title/tt0708942/quotes

In this program of research, we tested the hypothesis that *stereotypes* can create correspondingly relevant links when reasoning about power and gender. That is, we examined whether abstract stereotypes, when applied to a social category associated with power, can trigger concrete spatial processing of that category. This can be achieved only if the stereotype can interact with the space-power metaphor. As identified in the literature (see Fiske & Taylor, 2010, for an overview; Gilbert & Hixon 1991), the primary function of stereotyping is to simplify the complexity of the social world, which allows for efficient processing of social categorizations. In our research, we wish to extend this line of inquiry by demonstrating that stereotyping can be supported by another mental strategy associated with metaphoric links between abstraction and concreteness. This would imply that people employ two mental shortcuts simultaneously to process social stimuli to reduce the complexity of the relevant information.

Gender stereotypes associated with social power create expectations of how each gender should be evaluated in terms of power attributes (see Ellemers, 2018). In this light, we propose that stereotypic associations (men-powerful/women-powerless) elicit a strong link between gender and power and in this way evoke spatial processing via the space-power metaphor. That is, stereotypes would initiate spatial simulation of concepts that are otherwise not necessarily simulated spatially. Overall, we tested whether gender is simulated spatially when links between gender and power are present. Such links, as we assume, may lead to a semantic conflation between gender and power in the sense of *conceptual blending* (Fauconnier & Turner, 1998), which subsequently links gender with the spatial features associated with the power metaphor (i.e., *power is up*). In the present research, we combined the framework of conceptual blending with research on stereotyping and metaphors to report a novel demonstration of stereotypes functioning as a mental shortcut. This occurs through interactions with other mental tools: metaphors.

Conceptual blending

The conceptual blending perspective is relevant in examining how interactions between multiple mental representations occur. The theory refers to a process whereby semantics of two or more *consistent* concepts are combined (i.e., conflated). This process results in a *combined* meaning of the blended concepts. Previous research illustrates this process in the context of associations among primacy, left horizontal position, and goodness, in which semantics of these concepts are combined. Casasanto (2009, p. 362) stated: "Linguistic expressions like 'the prime example' conflate primacy with goodness (i.e., this phrase can mean the first example, the best example, or both). Speakers of languages like English may be predisposed to consider the leftmost item to be the first and therefore the best. This metaphorical blend of left, first, and best should result in a culturally constructed Good Is Left bias." This is in line with other research demonstrating that dominance (in the sense of abstract magnitudes, e.g., "older than," "richer than") is blended with primacy and hence associated with the left side in cultures with left-to-right writing/reading habits, whilst with the right side in cultures with right-to-left habits, where primacy is associated with the right (von Hecker, Klauer, Wolf, & Fazilat-Pour, 2016). The conceptual blending perspective is further supported by findings demonstrating spatial bases for associations between agency and gender on the horizontal dimension and rightward trajectory (i.e., moving from left to right). People associate men with agency, which in turn is related with the rightward spatial trajectory. As a result, there is an association between men being considered as agentic and their rightward-facing profiles (Suitner, Maass, & Ronconi, 2017).

Concreteness and abstraction. The involvement of spatial processes with respect to abstract concepts is established in the literature (Giessner & Schubert, 2007; Meier & Robinson, 2004; Schubert, 2005; von Hecker, Klauer, & Sankaran, 2013; Zanolie et al., 2012). Importantly, when mentally representing abstract concepts such as social power, people *simulate*, that is, mentally position, the concepts in the associated spatial locations. For example, Schubert (2005) found that when participants were presented with two groups (on high in power, the other low in power) displayed vertically on a screen, participants were faster at selecting the powerful group when it appeared at the top than bottom, and faster at selecting the powerless group when it appeared at the bottom rather than top.

However, it is important to note that these spatial processes are influenced by context. Spatial processing of abstract concepts in upper or lower vertical locations takes place *only* when mentally represented concepts are relevant to the conditions under which the processing occurs (Lebois, Wilson-Mendenhall, & Barsalou, 2015). That is, power is likely to be simulated in space when people think about concepts that are semantically related to power (e.g., judging which person - *master or servant* - is the powerful one, Schubert, 2005). However, it is unlikely that a power-related spatial metaphor would be applied to social groups when the processing does not occur in the context of power (e.g., judging which word - *master or servant* - has more letters; see also Taylor, Lam, Chasteen, & Pratt, 2015).

In relation to the previous argument, extant literature has focused solely on spatial processing of generic concepts that have strong power-related implications. Such strong implications might be more readily associated with spatial locations when the context is meaningful (judging a target's power level). That is, in Schubert's (2005) research, a long-term memory connection between presented stimuli and power via the literal meaning of the stimulus (e.g., boss, servant) was likely responsible for spatial simulations of powerful-top and powerless-bottom. Yet, in real contexts of thought and conversation, concepts vary in the degree to which they are associated with power. Our research aimed to address this gap.

Many concepts are malleable with respect to their associations and implications. Whether or not the relation between individuals belonging to specific social categories (e.g., men/women) and power is simulated in vertical mental space may be more malleable and dependent on whether the categories are seen as standing in a particular power relation or not. If they do, we argue that this will be a case when spatial processing or simulation of power is conflated or *blended* with the social categories (i.e., gender in our research), such that the social categories will be also simulated spatially. The simulation will *only* take place if the context allows for the creation of a relevant link between the power construal and the reasoned social category notion (e.g., "men" when seen in a power-context).

Associations among power, gender and space. Power and gender blending might likely arise due to social beliefs and gender role expectations that are associated with the socio-political structure within Western societies. This structure is primarily manifest in the patriarchy, where men usually have higher social status and more economical/political power than women (Ellemers, 2018; Guinote, 2017; Rudman & Phelan, 2010; Wood & Eagly, 2002). More generally, people are more likely to attribute stronger, socially powerful characteristics to men and weaker, powerless characteristics to women (DeWall & Maner, 2008; Eagly, Makhijani, & Klonsky, 1992; Eagly & Steffens, 1984; Rudman & Kilianski, 2000; Rudman, Moss-Racusin, Phelan & Nauts, 2012).

Building on and integrating the literature supporting a link between gender and power (e.g., Ellemers, 2018; Guinote, 2017), and power with space (e.g., Schubert, 2005; von Hecker et al., 2013), we hypothesize that the extent to which a person links gender with stereotypic attributes is related to spatial simulations of men at the top and women at the bottom. The underlying mechanisms of such simulation would be associated with the power-gender *blending* associated with stereotypic beliefs, whereby gender would acquire vertical correlates of power. Therefore, gender would be simulated spatially. Moreover, if stereotypes evoke spatial processing of gender on the vertical dimension via the blending of gender and power, then stronger stereotypes should involve more pronounced spatial processing. That is, holding a stronger versus weaker stereotype between gender and power would be associated with higher readiness to make stereotype-consistent categorizations in metaphorically congruent spatial positions (powerful-men-top; powerless-women-bottom). This is due to individual differences in mappings between concrete and abstract concepts. For example, spatial processing of the abstract concept of dominance in the upper vertical location is moderated by the extent to which people feel dominant (Robinson, Zabelina, Ode, & Moeller, 2008; see also Sherman & Clore, 2009). If we found that individuals who tend to more strongly associate men with power might also exhibit more pronounced spatial simulation of powerful-men at top versus women, this would demonstrate that individual differences in stereotype-consistent expectations might strengthen the link between two cognitive approaches occurring in social categorizations. Specifically, it would suggest that that having a higher tendency to use one mental shortcut, such as stereotyping, is likely associated with applying another one involving metaphors.

Alternatively, it may be the case that power, gender and space may not be represented by a blended simulation, but as simultaneously activated, non-interacting constructs. For example, multiple metaphorical mappings (valence-time, time-space, and space-valence) interact in temporal and valence judgements about verbs presented in the past/future forms (Spatola et al., 2018). People show a facilitation while responding to all metaphorically congruent stimuli (past-negative/future-positive verbs, past-left/future-right verbs, and left-negative/right-positive verbs). Although these three mappings (valence-time, time-space, and space-valence) are activated simultaneously, such representations are independent of each other suggesting that people globally process multiple mental metaphors.

If the relationship between the gender-power stereotype and the space-power metaphor is the same as among multiple metaphors, then in our studies, stereotypes and metaphors would be also activated independently. That is, participants would represent both male and female targets at the top when considering them as powerful. The opposite would be true for categorizing powerless individuals – both male and female targets would be represented at the bottom. Such findings would also corroborate Schubert's results (2005), further demonstrating that the space-power metaphor is stable and present regardless of the stereotype context. In terms of the stereotype activation, participants would be quicker to make stereotype-consistent judgements (attributing power to men and lack of power to women) as opposed to stereotype-inconsistent judgements (attributing power to women and lack of power to men).

Overview of studies

Integrating work on conceptual blending (Fauconnier & Turner, 1998), the stereotypic link between gender and power (Ellemers, 2018; Rudman & Kilianski, 2000) and metaphoric connections between power and verticality (Schubert, 2005; von Hecker et al., 2013), we suggest that stereotypic links between *men* and *powerful* as well as between *women* and *powerless* (i.e., stereotype-fit) should be associated with pronounced power-related spatial simulations, due to the interaction between the gender stereotype and the space-power metaphor. Such an interaction would be associated with *blending* between power and gender and in this way power-gender concepts would be linked with the spatial-metaphoric characteristics. Specifically, stereotypic categorizations would be present when people consider men as powerful over women and such thinking would involve spatial simulation of men at the top. Stereotypic categorizations would be also present when people consider women as powerless relative to men and this would involve spatial simulation of women at the bottom. Overall, when stereotypic categorizations are applied to making judgements in the context of power, then people should be more likely to expect men to be powerful and women to be powerless regardless of targets' spatial positions. Further, participants who hold strong stereotype-consistent expectations about attributes of each gender in the form of implicit associations between gender and stereotypic attributes (social status/emotionality) should exhibit pronounced spatial simulations while reasoning in stereotype-consistent ways.

To begin to test these predictions, we conducted six behavioral studies that adopted the same basic experimental paradigm. In each study, we asked participants to detect as fast as possible which target person (one female name; one male name) among two vertically displayed on a computer screen (top and bottom) was socially powerful or powerless. We measured the proportion of times participants' chose each gendered target (women/men) in each vertical position (top/bottom). Of greater importance, we also measured their reaction times to these choices. Although we slightly modified the procedure of each experiment (to assess the nature of the effects across different experimental conditions), our hypotheses were the same across all studies. We predicted that men considered as powerful would be categorized faster when their names were presented at the top than bottom. The opposite would be true for women considered as powerless, that is, we predicted that they would be categorized faster when presented at the bottom versus top. Also, we predicted that men would be categorized faster as being powerful when presented at the top in comparison to women considered as powerful when presented at the top. The opposite would be true for women – they would be categorized as powerless faster at the bottom than men as powerless at the bottom. We also examined the proportions of participants' choices of men and women as powerful or powerless, predicting that participants would be more likely to indicate men as powerful and women as powerless, due to stereotypic associations. We expected no spatial processing in the case of counter-stereotypic choices (i.e., selecting women as powerful or men as powerless). We hypothesized that holding stereotype-consistent expectations of gender in terms of men-high status, men-rationality, and top-rationality associations would likely involve quicker responses in selecting men as powerful at the top and women as powerless at the bottom (Rudman & Kilianski, 2000).

Finally, we tested whether the hypothesized associations among power, gender, and verticality were driven by people's direct associations between gender and verticality. Given that men are, on average, taller than women (Barhum, 2018), people might associate men directly with upper location due to men's height. To address this alternate hypothesis, we conducted a follow-up study to test if people associate gender with verticality when the power context is not relevant. We asked participants to quickly categorize gendered names as male or female. We manipulated the vertical location of the names by presenting them at the top or bottom of the screen (one name at a time). We measured how quickly participants categorized the names in each vertical position (we report the results of this study in a footnote in the results section).[2] As we believe that gender is associated with power via stereotypes and conceptual blending, we hypothesized that gender would not be associated with verticality directly.

Next, building upon the results of Studies 1-6, Study 7 assessed the interaction between gender stereotypes and the space-power metaphor with a physiological methodological paradigm involving pupillometry. We chose this measure as pupillary reactivity represents a relatively direct assessment of the stereotypic expectations, independent of behavioral responses. A large literature indicates that pupil size changes in response to expectancy-violating or conflict-inducing tasks. Specifically, increased pupil dilations are consistently observed early within trials presenting stimuli that are not consistent with people's expectations. For example, presenting anomalous playing cards (the king of hearts presented in black instead of red), incongruent Stroop and Simon task trials, or Thatcherized faces is sufficient to evoke increased pupil dilations as compared to trials conforming to people's expectations (see e.g., Lin, Saunders, Hutcherson, & Inzlicht, 2018;

[2] The methodological details of this study are presented in the Supplemental Online Materials. The data and materials are publicly available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/NEQH3).

Preuschoff, Hart, & Einhäuser, 2011; Proulx, Sleegers, & Tritt, 2017; Sleegers, Proulx, & van Beest, 2015; van Steenbergen & Band, 2013). Moreover, pupil dilation occurs in the case of behavioral response errors, independently of a conscious awareness that an error has occurred (Smallwood et al. 2011).

Given that widely held gender stereotypes have an implicit influence on people's expectations of typical male and female characteristics (Ellemers, 2018), we argue that these expectations can be directly tracked by pupil diameter. Further, if the gender-power stereotype interacts with space-power metaphor (forming expectations of powerful individuals being represented at the top), such a gender-power-space blend would create expectations of men considered as powerful at the top. Therefore, in Study 7, to test the validity and relevance of pupillometry to research on spatial simulations, first, we replicated Schubert's (2005) spatial task. We asked participants to quickly categorize powerful individuals on the screen while presenting them with vertically positioned (top versus bottom) power-related words (e.g., servant, master; adapted from Schubert, 2005) in one block. Second, in a separate block, we adapted the task used in Study 1 whereby participants were presented with vertically positioned gendered names and asked to quickly categorize the powerful person. We predicted increased pupil dilation on trials that represented stimuli in *incongruent* spatial locations (e.g., female names merely presented at the top or categorized as powerful at the top, and powerless groups, e.g., servants, presented at the top).

Studies 1 – 6: Investigating the stereotype-metaphor blend

The primary aim of Studies 1-6 was to investigate the stereotype-metaphor blend in the context of gender, power and verticality relations. Across the studies, we introduced minor variations in our paradigm to test whether people are equally likely to combine metaphors and stereotypes under different task conditions. Our rationale was derived from research by Lebois et al. (2015), who demonstrated that people rely upon spatial cues only when those cues are salient and relevant to the experimental task. This suggests that the context under which people make judgements is important in determining the potential involvement of spatial associations in social decision-making.

Therefore, in Studies 1 and 2, we tested whether presenting participants with ambiguous task conditions (i.e., providing only gendered names) would be associated with the most optimal conditions for participants to combine metaphors with stereotypes. In Studies 3 and 4 we tested whether introducing less ambiguity (by pairing gendered names and gender-neutral, high-status professions) would be associated with a tendency to use one mental strategy (e.g., using only metaphoric mappings between space and power, therefore, a tendency to choose targets faster at the top regardless of their gender). Across these four studies, we used a status IAT in keeping with Rudman and Kilianski (2000), who showed that people have negative attitudes towards women in power positions due to their implicit associations between men and status. Stereotypically, people perceive men with high power and a high social status. Therefore, we wanted to address whether people's stereotypic associations between high-status and men would be also involved in the mental representation of metaphoric and stereotypic thinking associated with power.

In Studies 5 and 6, we considered whether more ambiguous task conditions (as in Studies 1 and 2) would interact with participants' stereotypes about gender in terms of rationality and emotionality. This follows from Cian et al. (2015), who found that rationality is associated with upper vertical positions, whilst emotionality is associated with lower vertical positions. Stereotypically women are associated with emotionality, whilst men with rationality (e.g., Hess et al., 1999). We examined whether people who endorse such stereotypic views in combination with emotionality-bottom/rationality top associations would be more likely to rely on stereotypes over metaphors in attributing power to gender.

To achieve the above aims, we asked British (Study 1) and Polish (Study 2) participants to quickly detect either the powerful (condition 1) or the powerless (condition 2) person on the basis of gendered names only (male or female) that were vertically displayed (top or bottom) on a computer screen. Next, in Study 3, we paired both male and female names with high-status, gender-neutral professions to promote power-related thinking in a professional context. In Study 4, we primed participants with either social status represented as professions or neutral words. Across Studies 1-4, after presenting the spatial task, we asked participants to complete an IAT measuring their associations between status and gender. In the final two behavioral experiments we investigated the same assumptions as in Studies 1 and 2, but additionally we examined potential correlates of the investigated spatial simulations by measuring participants' stereotypic associations: men-high status; men-rationality; top-rationality (Studies 5 and 6).

Table 1 summarizes the manipulated variables. Across all studies, we measured how likely participants were to categorize male and female names as powerful/powerless, as well as the reaction times to their choices. First, we present an overview of the methodology used across all six studies and a combined analysis. The method sections and a summary of results of each individual study are available in Appendix B and F, respectively. From there, we present the methodological details and results of Study 7.

Participants

Demographic information about participants is presented in Table 2. The sample size of individual studies was informed by a power analysis based on relevant previous research (e.g., Schubert, 2005; von Hecker et al., 2013, for Cohen's $dz = .30$). The power analysis was estimated in G*Power software for a three-way interaction involving effects based on an analysis of variance design to achieve .80 of power, at an $\alpha = .05$; Faul, Erdfelder, Buchner, & Lang, 2009). The analysis estimated between 46-62 participants per study. Upon completion of the studies, we computed a post-hoc power analysis for ANOVA-based linear mixed models using PANGEA application (Westfall, 2016). For the obtained effect sizes (ranging from dz = .30-.33), the analysis revealed that we achieved .95 power. Participants for five of the studies were recruited from a sample of Cardiff University students and received either course credit or payment for their participation. Participants for the remaining study were recruited from Maria Curie-Sklodowska University of Lublin, Poland, and received course credit. Across six experiments, we recruited 379 participants in total (331 females; mean age = 19.79 years).[3] All studies were approved by the Ethics Committee at Cardiff University (approval number: EC.16.03.08.4484GA).

Exclusions: One participant was excluded from Study 1 and 4 each, and two participants were excluded from Study 3 and 5 each due to extreme response times, as determined by Tukey criterion (specifically, their response times were larger than the upper quartile plus 3 times the interquartile range; Clark-Carter, 2004).

[3] The high proportion of female participants in our studies is attributable to the demographic makeup of the participant panels.

Table 1

Overview of the manipulated factors (Studies 1 - 6)

Factor	Levels	Design
Task	Powerful versus powerless	Between-participants
Trial type	Men-top versus women-top	Within-participants
Gender choice	Men versus women	Within-participants/quasi-experimental variable

Table 2

Demographic information about participants.

Studies	N	Gender	Mean Age	Nationality
Study 1	78	66 women	20.17 years	British
Study 2	73	58 women	20.19 years	Polish
Study 3	51	47 women	19.04 years	British
Study 4	58	51 women	18.97 years	British
Study 5	59	57 women	22.12 years	British[4]
Study 6	60	52 women	21.40 years	British

[4] 26% of the sample were non-native English speakers, but were fluent in English.

Materials

All Studies. Spatial Task. We selected 10 popular British (Studies 1, 3, 4, 5, and 6) and Polish (Study 2) names (5 female and 5 male).[5] Each name was randomly paired with a name of the opposite gender (e.g., Oliver – Emily, see Appendix C). Across all studies, the matched pairs were then presented on the computer screen in white letters on a black screen (font size 15 – 21 across the studies). In Study 3, we additionally selected professions (derived from a pilot study)[6] that were assigned to male and female names (adapted from Study 1; e.g., Oliver-Professor, Sophie-Professor). Each pair (profession and name) was assigned to a corresponding pair including a name of the opposite gender. We created all combinations of professions and gender (e.g., male scientist - female professor; female professor - male scientist) within four sets of pairings (see Appendix D). Participants were randomly assigned to receive one set of combinations. In Study 4, we introduced another minor modification. Before each trial, (in which they were required to indicate the powerful/powerless person at the top or at the bottom) participants were presented with a social status item (gender-neutral profession adapted from Study 3, i.e.: scientist, architect, doctor, professor, dentist) or neutral word (i.e., vegetables: carrot, potato, lettuce, broccoli, cabbage). The prime was manipulated within-participants. After making decisions about the powerful/powerless person in each trial, participants reported whether the initially presented prime word belonged to the social status category (by pressing the arrow pointing left), or to

[5] See the most popular names: Popular British baby names: Year by year. (n.d.). Retrieved February 9, 2015, from http://www.babycentre.co.uk/popular- baby-names (Study 1); Academy of childbirth. (n.d.). Retrieved May 2, 2016, from: http://akademiaporodu.pl/top-news/najpopularniejsze-imiona-w-2015-2016-mapy-ranking (Study 2).

[6] In a pilot study, 30 participants rated 30 professions on levels of power, agency, gender-typicality and social status (0: low; 10: high). The mean ratings revealed five professions (doctor, dentist, architect, scientist, professor) with high power (M = 6.92; SD = .85), social status (M = 8.20; SD = .83), being gender-neutral (M = 4.34; SD = .53) and agency-neutral (M = 4.68; SD = .83).

the vegetable category (by pressing the arrow pointing right).[7] In this way, the prime word remained activated during participants' decision about the powerful/powerless person.

Design

All Studies. Spatial Task. The same basic design was used in all six studies. Task ("find powerful" versus "find powerless") was always manipulated between-participants while the trial type (male name/top and female name/bottom versus female male/top and male name/bottom) was always manipulated within-participants. There were four blocks of trials. Each block included a presentation of five pairs shown twice (ten total trials). In half of the trials, a male name (e.g., Oliver) was displayed at the top of the screen with a female name (e.g., Emily) at the bottom (see Figure 1). Both names were centered and there was a 23cm vertical distance between the names. In the remaining trials, this display was reversed (women at the top and men at the bottom). There were 40 trials in total. The order of trials within each block was randomized. The same design was applied to Study 3, except that instead of presenting pairs of gendered names, we presented gendered names paired with professions (see Figure 2). In Study 4, prior to presenting the first half of the trials (men-top; women-bottom) participants were primed with a social status item, and prior to presenting the other half (men-top; women-bottom) they were shown a neutral word (the prime words were selected at random from a list of vegetables; see Figure 3). The same was done for the trials where women were presented at the top and men at the bottom, so there were 20 trials within each block (80 trials in total). The trials were also randomized within each block within-participants; there were four blocks in total. We measured how quickly participants responded and which gender they picked as powerful/powerless on each trial type (see Table 1 for an overview of the manipulated variables).

[7] The responses were reversed for the other half of our participants: they pressed the arrow pointing right for social status items and the arrow pointing left for vegetables (participants were randomly assigned to one of the key arrangements).

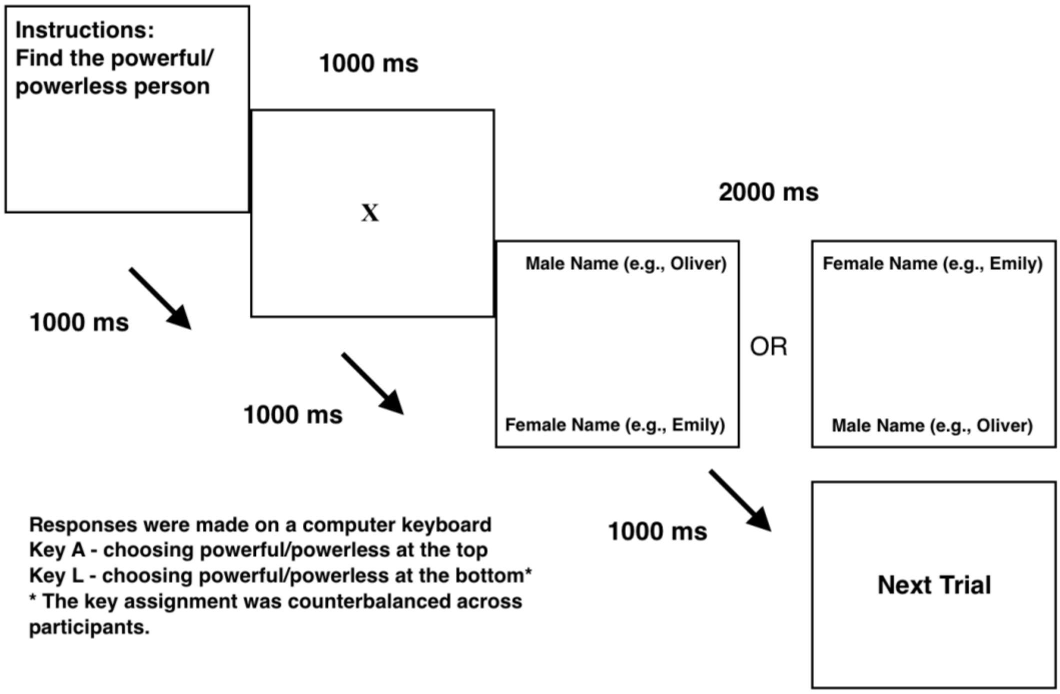


Figure 1. Spatial task procedure of a single trial (Studies 1, 2, 5 and 6).[8]

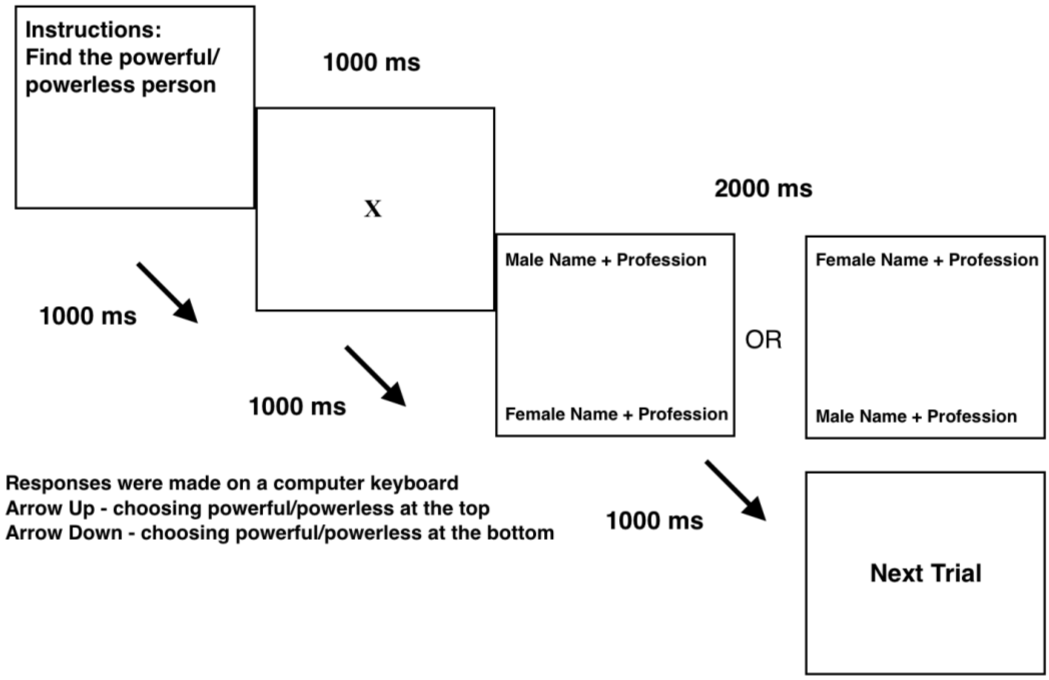


Figure 2. Spatial task procedure of a single trial (Study 3).

[8] In Study 5, participants used the arrow pointing up to indicate the powerful/powerless at the top and the arrow pointing down to indicate the target at the bottom.

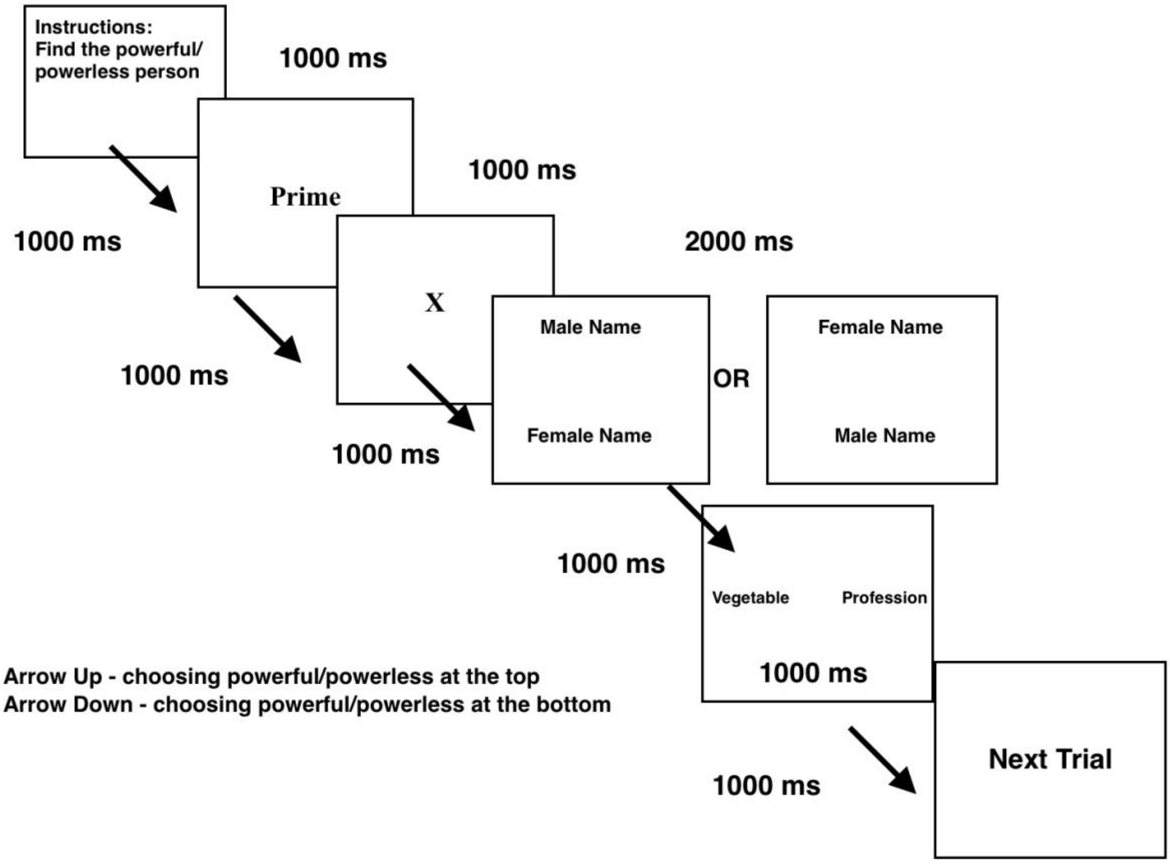


Figure 3. Spatial task procedure of a single trial (Study 4).

Studies 1, 2, 3, and 4. Status-gender IAT. We adapted the standard IAT of implicit attitudes towards high- or low-status men and women from research by Rudman and Kilianski (2000; see Appendix E). Within a single block, participants were asked to categorize items that belonged to either high status (boss, supervisor, expert, leader, executive, authority) and men (e.g., Paul, Brian, Kevin) or low status (secretary, helper, aide, clerk, subordinate, assistant) and women (e.g., Lauren, Kate, Zoe), or vice versa in another block.[9] The order of the blocks was counterbalanced across participants. In Study 2, the instructions and categories were presented in Polish. In Study 3, we reduced the number of blocks of a standard IAT from five to two (see Sriram & Greenwald, 2009). Participants

[9] In the current paper, we disregard conceptual independence between power and status, and we focus on the conceptual overlap (see von Hecker et al., 2013 who found associations between verticality and status).

were asked to keep in mind two target categories within two blocks (men and high status in one block or women and high status in another) and respond as fast as possible by pressing a designated key when they saw an item that belonged to the target category on the screen. When they saw distractors (women and low status or men and low status) they were to press a second designated key.

Studies 5 and 6. Rationality/emotionality-gender IAT. We asked participants to quickly and accurately sort items that belonged to the category of emotionality and women as well as rationality and men, and vice versa. The items are presented in Appendix E.

Studies 5 and 6. Rationality/emotionality-verticality IAT. Participants categorized items that belonged to the category of up and rationality, and the category of down and emotionality, or up/emotionality and down/rationality (the items for emotionality/rationality categories were the same as in the rationality/emotionality and gender IAT). Verticality items are presented in Appendix E. Both IATs in Studies 5 and 6 were designed in the same standard way.

All Studies. Procedure.

All studies were presented using DirectRT (Jarvis, 2012). Participants sat approximately 70cm away from the computer screen. After providing consent, participants read the instructions and received verbal definitions of *socially powerful* and *socially powerless* individuals (adapted from Galinsky, Gruenfeld, & Magee, 2003). In the powerful condition, we provided the following definition: 'By powerful, we mean a person that controls the ability of another person to get something they want, or is in a position to evaluate that person', whilst in the powerless condition we stated: 'By powerless, we mean a person whose ability to get something they want is controlled by another person, and the other person is in a position to evaluate them.' In Study 2, as there are no direct semantic equivalents of the adjectives *powerful* and *powerless* in Polish, we used substitutes. As social power can be defined in terms of social influence (Lewin, 1941), we asked Polish participants to find *influential* or *not influential* individual on the screen. Then, participants completed the spatial task by pressing keys on the computer keyboard to indicate the powerful/powerless person on the screen. After the spatial task, all participants completed the IATs (the status-gender IAT in Studies 1-4; rationality/emotionality-gender and –verticality IATs in Studies 5 and 6). At the end of each session participants were debriefed. Each experiment lasted approximately 20 minutes.

Analytic strategy

To provide an overall picture of the expected spatial effects, we analyzed the data from all six behavioral experiments by estimating a hierarchical linear model, in which we nested *participants* within a contextual variable - our *studies*. Before conducting the analysis, we assessed the appropriate random structure (i.e., random slopes) that would best fit our data and then we estimated a final hierarchical linear model to evaluate the fixed effects, that is, gender choice (men versus women) and trial type (men-top versus women-top; see Jaeger, 2008; Judd, Westfall, & Kenny, 2012; see Table 1 for the factors involved). Fitting random slopes is associated with maximizing power because all variance that is modelled as interactions between participants and the levels of a given experimental factor (and used as a predictor) would conventionally be attributed to measurement error. It is the power-maximizing advantage of linear mixed models that both fixed and random factor are introduced on the predictor side of the model. The way of determining the best random structure for the overall analysis is presented in Appendix A. The same strategy was applied to the data collected in Study 7. The effect sizes are reported as Cohen's *dz*.[10] Finally, we wish to note that in our experiments the predictor variable associated with gender choice is quasi-experimental.

[10] We computed Cohen's *dz* using the following formulae: t-value divided by the square root of the sample size. The relevant t-values and sample sizes are reported in the Results sections.

Results across six studies

Proportions of choices. First, we analyzed the proportions of participants' choices. As predicted, participants were more likely to choose men as powerful (59.5%) rather than women as powerful (40.5%), $\chi^2(1) = 658.78$, $p < .001$, $dz > 1.42$, odds ratio $_{\text{choosing men as powerful versus women}} = 1.48$ (95%CI[1.43, 1.53]). We also tested whether proportions of choices were correlated with reaction-times, but the analyses indicated that these two processes were independent, *ps* > .13.

Response latencies. In terms of reaction times of participants' choices, our data were hierarchical, so we built a multilevel linear model in which intercepts varied across *participants* that were nested within *studies.* Next, assessing the appropriate random structure that would best fit our data revealed that the model fit improved after including *gender choice* as a random slope (see Appendix A).[11] We then estimated the final model to evaluate the fixed effects - that is, gender choice (men versus women),[12] trial type (men-top versus women-top), and task instructions (categorize powerful versus powerless). We corrected for multiple comparisons by applying the Tukey adjustment.[13]

As predicted by the conceptual blending account, the three-way interaction among gender choice, trial type, and task was significant, $F(1, 16113) = 5.81$, $p < .016$, $n = 379$ (see Figure 4). Starting with the powerful condition, as hypothesized, participants were significantly faster at selecting men as powerful when they were presented at the top of the

11 A step-wise removal of random factors within the model (*gender choice* and *participants*) provided the same results estimation as compared to a model involving the random factors.

12 As the nature of our results is correlational, we also ran a model in which we predicted *gender choice* from task, trial type, and reaction time. As expected, a mixed-effects logistic regression indicated that the interaction among trial type, task, and reaction time was significant in predicting gender choice, $\beta = -.27$, 95% CI[-.47, -.07], SE = .10, $z = -2.71$, $p < .001$. After adjusting the significance level ($\alpha = .01$) to account for multiple comparisons, we found that a faster reaction time was associated with a higher probability of selecting male names than female names only in the powerful condition and when men were presented at the top and women at the bottom, $\beta = -.29$ 95% CI [-.39 , -.19], SE = .05, $z = -5.59$, $p < .001$. The other post-hoc tests did not show significant associations between reaction time and gender choice $p > .04$.

13 The same results were obtained with Bonferroni and Holm adjustments.

screen (M = 1019 ms; SE = 24.16, 95% CI[986, 1081) compared to selecting women as powerful when they were at the top of the screen (M = 1107 ms; SE = 25.78, 95% CI[1038, 1140]), $t(597) = 4.38$, $p < .001$, $dz = .33$ (95% CI[.18, .48]; n = 178). Interestingly, in the powerless condition, participants did not differ in how quickly they selected men (M = 1108 ms; SE = 24.66, 95% CI[1085, 1182]) versus women (M = 1057 ms; SE = 25.51, 95% CI[1058, 1159]) as powerless when they appeared at the bottom, $t(733) = .85$, $p = .59$, dz =.07, (95% CI[-.09, .22]; n = 169). Importantly, the contextual variable, *studies*, did not moderate the three-way interaction, $F(5, 16049) = 1.19$, $p = .311$. The above analyses indicate that when power (but not the lack of power) was conceptually combined with gender in a stereotype-consistent manner, participants spatially simulated men on the top. However, this was not the case when a woman was selected as powerful (i.e., in the context of a stereotype-inconsistent choice).[14]

Subsequently, to determine whether the above effects were specifically associated with spatial locations, we conducted additional post-hoc analyses, whereby we divided our dataset according to *stereotype-consistent* (i.e., when participants selected men as powerful and women as powerless) and *-inconsistent* choices (i.e., when participants selected men as powerless and women as powerful). Consistent with our expectations, these analyses revealed that men were selected as powerful faster (the stereotype-consistent choice) when they appeared at the top of the screen (M = 1019 ms; SE = 9.34, 95% CI[991, 1088]) versus the bottom of the screen (M = 1072; SE = 24.65, 95% CI[1035, 1132]), $t(9530) = 4.72$, $p <$ .001, $dz = .35$ (95% CI[.20, .50]; n = 184). The reverse was true for choices of women (the stereotype-inconsistent choice) – there was a non-significant pattern whereby women were

14 The results of our follow-up study (designed to account for the possibility that top-men-powerful response advantage was due to the direct association between gender and verticality) indicated that gender did not interact with location of the presented stimuli, $F(1, 4161) = .15$, $p = .700$, see Supplemental Online Materials for detailed results. This demonstrates that gender is not directly associated with verticality in the absence of the power context.

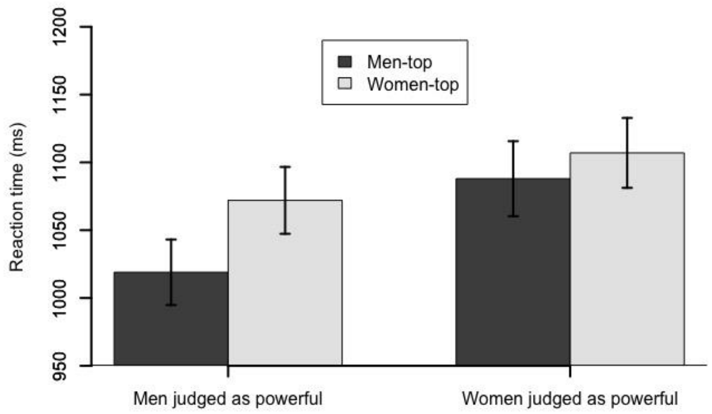
selected as powerful more slowly when they appeared at the top of the screen (*M* = 1107 ms; *SE* = 27.04, 95% CI[1049, 1156]) versus the bottom of the screen (*M* = 1088 ms; *SE* = 27.63, 95% CI[1070, 1178]), *t*(6329) = 1.66, *p* = .097, *dz* = .13 (95% CI[-.03, .29]; *n* = 158).

In contrast to our hypothesis, in the powerless condition, female names were not selected significantly slower when they appeared at the top of the screen (the stereotype-consistent choice; *M* = 1102 ms; *SE* = 24.48, 95% CI[1072, 1168]) versus the bottom, (*M* = 1057 ms; *SE* = 24.77, 95% CI[1067, 1165]), *t*(9648) = .46, *p* = .648, *dz* = .03 (95% CI[-.12, .18]; *n* = 169). This difference was also not significant when men were selected as powerless (the stereotype-inconsistent choice) when they appeared at the top of the screen (*M* = 1114 ms; *SE* = 27.11, 95% CI[1081, 1187]) versus the bottom (*M* = 1108 ms; *SE* = 26.98, 95% CI[1098, 1204]), *t*(6298) = 1.36, *p* = .173, *dz* = .10 (95% CI[-.05, .25]; *n* = 175).[15] Overall, the three-way interaction suggests that the top-advantage is present in reaction times when men are selected as powerful, supporting our hypotheses regarding the selection of stereotypically powerful targets.[16]

15 The same results were obtained when stereotype-consistent and -inconsistent responses were combined in one analysis.

16 See Online Supplemental Materials for further analyses of main effects and two-way interactions.

(a) Powerful condition



(b) Powerless condition

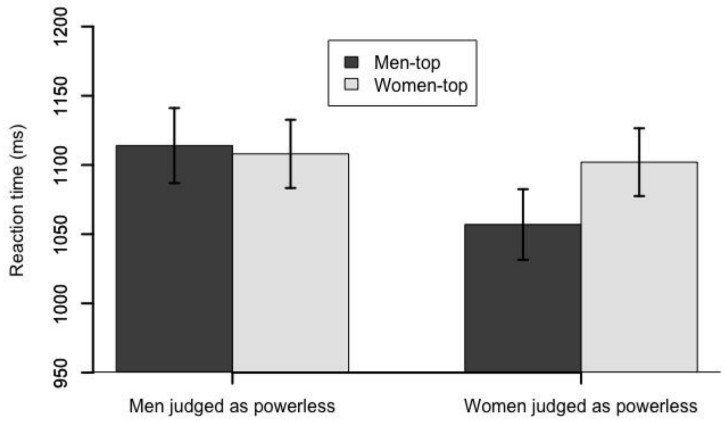


Figure 4. Response latencies to (a) men and women judged as powerful, and (b) men and women judged as powerless according to the trial type (men-top/women-bottom or women-top/men-bottom) across six experiments. The error bars show -/+1 standard error.[17]

[17] Note that "Men-top" refers to trials where male names were presented at the top and female names at the bottom, whilst "Women-top" represents trials where female names were presented at the top and male names at the bottom.

High status-gender IAT. To test whether the above findings were associated with individual differences in implicit associations between gender and social status, we analyzed participants' IAT responses. After controlling for the relevant studies (1- 4), we found that participants were significantly faster at categorizing men and high status items (M = 794 ms; SE = 12.54, 95% CI[769, 818]) compared to women and high status items (M = 882 ms; SE = 10.52, 95% CI[799, 841), $F(1, 256) = 5.25$, $p < .023$, $\eta_{2p} = .02$, (95% CI[.01, .06]). The studies by IAT category (men-high status; women-high status) interaction was not significant, indicating that the direction of the effect of the category did not differ across studies, $F(1, 256) = 2.54$, $p = .112$, $\eta_{2p} = .01$ (95% CI[.01, .04]). Second, we calculated a facilitation score for each participant. This was done by subtracting RTs from blocks in which participants responded to the category combination of men/high status from blocks in which they responded to women/high status; such that positive scores indicated a tendency to associate men with high status. Linking this index with our RT data, we also computed two separate difference scores subtracting the RTs for choices of men at the top from the RTs for choices of men at the bottom and, vice versa, the RTs for choices of women at the bottom from the RTs for choices of women at the top. The more positive these scores are, the more impactful is the simulation of men as powerful at the top. Therefore, we term these scores "simulation scores." After controlling for studies, we found that the IAT facilitation scores (i.e., the tendency to associate men with high status)[18] did not correlate with the spatial simulation of men as powerful at the top, $r(128) = -.07$, 95% CI[-.25, .10], $p = .434$. The same was true for the simulation of women as powerless at the bottom, $r(108) = -.06$, 95% CI[-.23, .12], $p = .528$. Likewise, the correlation between facilitation score to men as powerful at the top

[18] we also correlated the absolute scores, i.e., participants' reaction times to choices of men as powerful on the top and women as powerless on the bottom, with the IATs facilitation scores, and we found similar results as when correlating the simulation scores.

versus women as powerful in the same spatial location and IAT facilitation score was not significant, $r(104) = .06$, 95% CI[-.13, .25], $p = .621$.

Rationality/emotionality-gender IAT. After controlling for Studies 5 and 6, we found that participants were more likely to associate men with rationality ($M = 697$ ms; $SE = 9.41$, 95% CI[678, 865) compared to women ($M = 840$ ms; $SE = 12.48$, 95% CI[816, 865), $F(1, 117) = 4.25$, $p < .041$, $\eta_{2p} = .04$ (95% CI[.01, .10]). The choice category by studies interaction was not significant, $F(1, 117) = .56$, $p = .457$, $\eta_{2p} = .01$ (95% CI[.01, .05]). Similarly, to the status-gender IAT, we calculated facilitations scores – the higher the score, the higher the tendency to associated men with rationality. Such associations did not significantly correlate with spatial simulations scores of men as powerful at the top, $r(58) = .11$, 95% CI[-.15, .36], $p = .427$, or women as powerless at the bottom versus top, $r(57) = .18$ 95% CI[-.08, .42], $p = .180$. The same was true for the correlation between the facilitation score to men as powerful at the top versus women as powerful at the top and IAT facilitation score, $r(51) = .16$, 95% CI[-.07, .46], $p = .264$.

Rationality/emotionality-verticality IAT. We found that participants did not associate top vertical positions more with rationality ($M = 784$ ms; $SE = 13.68$, 95% CI[757, 812) than with emotionality ($M = 842$ ms; $SE = 15.48$, 95% CI[811, 873), $F(1, 114) = 1.50$, $p = .223$, $\eta_{2p} = .01$ (95% CI [.01, .07]; after controlling for Studies 5 and 6). The interaction between the IAT category and studies was also not significant, $F(1, 114) = .75$, $p = .389$, $\eta_{2p} = .01$(95% CI [.01, .06]. Subsequently, we calculated an IAT facilitation score – the higher the score the higher the tendency to associate rationality with top positions. The associations did not correlate with the spatial simulation of men as powerful at the top versus bottom, $r(56) = .11$, 95% CI [-.36, .16]; $p = .426$, or of women as powerless at the bottom versus the top, $r(51) = -.13$, 95% CI [.-.39, .15], $p = .350$. The facilitation score to men as powerful versus women as powerful at the top was not correlated with the IAT facilitation scores, $r(49) = .33$, 95% CI [.05, .55] $p = .020$.[19]

Summary of results of Studies 1-6

The results of these six studies supported our hypotheses in the powerful condition. As predicted, we found spatial processing of gender when the gender-power stereotype interacted with the space-power metaphor. Specifically, attributing high power to the male target (i.e., a stereotypic categorization) involved spatial simulations, but that was not the case when high power was attributed to the female gender (i.e., a counter-stereotypic categorization). This was demonstrated by significantly faster categorizations of men than women as powerful when their names were presented at the top. Our data further indicated that the above effects involved simulations on the vertical dimension, as men as powerful were chosen faster at the top than at the bottom of the screen. Contrary to predictions, this pattern was not true for women selected as powerless.

Considering alternative perspectives, the findings are consistent with Schubert's (2005) results showing that thinking about power involves spatial processing – *powerful* being simulated at the top rather than bottom. However, this process did not occur when participants reasoned in counter-stereotypic ways (i.e., selecting a woman as powerful or a man as powerless). In light of Schubert's findings, whereby powerful groups were simulated at the top and powerless at the bottom, we would expect that considering women as powerful and men as powerless should also involve spatial correlates. Therefore, our findings suggest that spatial processing of power is involved when power is attributed to a target concept when this concept can be *blended* with power in socially meaningful ways – for example, by applying a stereotype. Thus, our work suggests that spatial processes involved in abstract

19 To account for multiple correlations, we reduced the alpha level to .01.

thinking about power are moderated by social thought, demonstrating that stereotypes interact with metaphors in social categorizations. In addition, we found that such simulations were independent of implicit attitudes towards gender and social status or rationality.

Study 7: Investigating spatial simulations with pupillometry

In Study 7, we further assessed the interaction between stereotypes and metaphors in gender categorizations by employing a physiological methodology of pupillometry. Studies 1-6 found behavioral evidence that gender stereotypes conceptually interact with the space-power metaphor in social power categorizations. Next, we wanted to examine if these associated concepts formed the basis for pre-conscious stereotype-spatial *expectations*. We measured these expectations with a physiological recording of pupil size. This procedure not only allows for measuring a physiological, pre-conscious response, it also allows us to address potential demand characteristics associated with the behavioral paradigm.

Research on changes in pupil size demonstrates that early surprise or expectancy violation evoked by the mere presentation of incongruent stimuli correlates with increased pupil dilation. The increased dilation is associated with heightened attentional arousal, which is linked with neural brain activity of locus-coeruleus norepinephrine system (LC-NE; see Aston-Jones & Cohen, 2005, for an overview). The LC-NE system activates in response to salient stimuli and modulates task engagement or withdrawal. When people are engaged with a task, the LC is activated in a phasic mode of firing that produces a rapid release of cortical noradrenaline (NE) to the parietal and frontal areas involved in cognitive control, for example, the anterior cingulate involved in processing task demands (Murphy, Robertson, Balsters, & O'Connel, 2011). Such release increases neuronal gain in the relevant brain areas aiding stimulus processing (Aston-Jones & Cohen, 2005). Consistent with this perspective, Oliveira, McDonald, and Goodman (2007) found that the anterior cingulate is active when people experience unexpected events such as receiving positive (negative) feedback when expecting negative (positive) feedback. Evidence from other studies also indicates that in general, arousal is increased when expectations are violated in social situations, e.g., due to social exclusion or interaction with someone who does not conform to social expectations (Mendes, Blascovich, Lickel, & Hunter, 2002; Moor, Crone, & van der Molen, 2010).

A more direct psychophysiological proxy for LC-NE response to expectancy violation is pupillary dilation (Gilzenrat, Nieuwenhuis, Jepma & Cohen, 2010). For example, Proulx et al. (2017) recorded participants' pupil size while presenting neutral, angry, as well as Thatcherized and upside-down faces, with the latter two categories representing expectancy-violating human faces. They found that both types of expectancy-violating faces evoked earlier and larger peak pupil dilations than neutral faces. These findings are consistent with pupillary response to non-social inconsistencies, such as viewing reverse-colored playing cards (e.g., a black Queen of Hearts, Sleegers et al. 2015), or inconsistent Stroop trials (see Laeng, Ørbo, Holmlund, & Miozzo, 2011; Rondeel, Van Steenbergen, Holland & van Knippenberg, 2015; see also Lin et al., 2018). Taken together, these findings emphasize that pupil size is associated with expectancy violation regardless of the type of perceived inconsistencies. Simple perceptual and semantic mismatches, as well as inherently social incongruencies associated with surprise evoke early, heighten pupillary dilations.

Pupil size in stereotype-metaphor expectations

Integrating the literature on pupillometry (e.g., Lin et al., 2018; Proulx et al., 2017; Sleegers et al. 2015), stereotypic expectations (Ellemers, 2018), as well as our previous findings indicating the stereotype-metaphor interaction in social judgements, we investigated whether the stereotype-metaphoric expectations could be detected at the level of pre-conscious processing *independent* of behavioral responses. As a secondary aim, we also investigated response-locked cognitive effort on trials where participants could potentially attempt to override the stereotype-metaphoric consistency. In a recent review, van der Wel and van Steenbergen (2018) suggested that cognitive effort associated with processing of inconsistent information can be also detected by pupil dilations, such that higher cognitive effort involves higher pupil dilations. Such dilations can be observed after a behavioral response is made in a task and might constitute a physiological measure of effort, which is complementary to the reaction-time methodology.

To achieve these aims, we designed an eye-tracking experiment involving two stages. In the first stage, we presented participants with Schubert's task (2005, Study 2), in order to test whether pupil size changes can indeed parallel the effects obtained using the reaction-time methodology applied by Schubert (2005). Because Studies 1-6 detected significant spatial effects in the powerful condition, we limited the task to selecting the powerful member of the pair. Participants were presented with manipulated vertical positions (top versus bottom) of generic powerful (*master*) and powerless groups (*servant*) and asked to detect as quickly as possible the powerful group on the screen. We hypothesized that presenting participants with powerful groups at the bottom of the screen versus the top would be associated with an initial increase in pupil size due to response-preparatory surprise. This would be observed in the *pre-response* period of a trial, that is, before participants indicate the powerful person (see Laeng et al., 2011, for a similar procedure). If people hold expectations associated with the space-power metaphor, then the violation of such expectations should be processed at a pre-conscious level of processing before participants make a behavioral response. However, having responded to an incongruent trial (powerless-bottom) in contrast to a congruent trial (powerful-top) should be associated with increased response-locked cognitive effort leading to continuous pupil dilation. This would be observed in the *post-response* period.

In the second component of the experiment, we presented participants with the spatial task used in Study 1. Again, we asked participants to select as quickly as possible the powerful person from a pair of gendered names. We also manipulated the trial type. On men-top trials, participants were presented with a male name presented at the top and a female name at the bottom and vice versa on women-top trials (see Figure 5). Participants' categorizations of male and female names as powerful in each vertical position were recorded and constituted a quasi-experimental variable. We predicted that participants would exhibit increased initial surprise associated with cognitive conflict, and thus increased pupil dilation, with the presentation of women-top trials as opposed to men-top trials in the *pre-response* period. They would then demonstrate further pupillary dilation associated with the response, only *after* indicating women as powerful at the top as opposed to men as powerful in the same location, i.e., in the *post-response* period. We speculate that such higher cognitive effort, as reflected by increased pupil dilation, should be observed in response to potential inhibition of the automatically activated existing association between powerful-men and verticality. Specifically, pupil dilation would indicate an attempt to control the automatic gender-power-space expectations. We also predicted that the presentation and selection of male names when they appeared at bottom versus top would be associated with increased pupil dilation. In our previous studies, we did not find differences in simulating women as powerful at the top versus the bottom, hence, we predicted no differences in cognitive conflict or effort in those cases.

Overall, across both response periods, pupil size should be increased on *incongruent* Schubert and gender trials. Both stages of the experiment, that is, the Schubert block and the gender block, were presented to all participants. Their order was counterbalanced across participants. Within each block, we measured participants' pupil size within each trial. To acquire enough pupil size data in the post-response period, we presented each trial for 4000 ms, however, we encouraged participants to respond as quickly as they could. In the gender block, we also recorded how often participants picked each gender in each vertical position.

In both blocks, each trial lasted 4000 ms. Therefore, if participants did not respond within that period, their answers were not recorded (such responses constituted less than 1% of the data). For parsimony, we present the analysis of Schubert block in the Online Supplemental Materials, while the gender block analysis is presented below. We summarize the results of both blocks in the discussion.

Method

Participants

Sixteen Cardiff University undergraduate students participated in the experiment (13 women, one non-binary, mean age = 19.44; the power analysis in G*Power software for a large effect size derived from Proulx et al., 2017, Cohen's dz = .65, based on an analysis of variance design to achieve .80 of power at an α = .05, estimated the required sample size of 14 participants; Faul et al., 2009). The PANGEA application estimated the achieved power of .75 for the obtained effect sizes (Cohen's dz = .45-.80; Westfall, 2016). Participants received course credit for taking part in the study.

Materials

For the gender block trials, we adapted the spatial task materials from Study 1. That is, we presented five pairs of gendered names as stimuli, such that on half of the trials male names were positioned at the top and female names at the bottom or vice-versa on the other half of the trials. In the Schubert block, we presented five pairs of powerful and powerless groups – we randomly selected five powerful and five powerless groups from the list of items used by Schubert (2005, Study 2). To standardize the conditions of both the gender and Schubert blocks, we also randomly paired each powerful group with a powerless group and presented the same pairs across the experiment (master-servant, coach-athlete, employer-employee, officer-soldier, and boss-secretary). The vertical position of the powerful-powerless pairs was also manipulated in the same ways as in the gender block. Each block included 40 trials, therefore, there were 80 trials in total.

To measure participants' eye movements and pupillary activity, we used a Tobii X3-120 screen-based and non-invasive eye tracker that samples data at the speed of 120Hz (Tobii, Stockholm, Sweden). The eye tracker was integrated in a 17'' computer screen with a resolution of 1920 x 1080 pixels. The pupil size output is corrected by an in-built algorithm that compensates for eye movements and changes in the angle at which pupil size is recorded.

Design

We used a within-participants design, so that each participant contributed to both the gender and Schubert blocks. We asked participants to select the powerful person on the screen. The order of blocks was counterbalanced between-participants and trials were randomly presented within each block. The trial type in gender block (men-top/women-top) and the vertical position of powerful and powerless groups (powerful-top/powerless-bottom) was manipulated within-participants. In the Schubert block, we measured participants' pupil size to the trials in which they made a correct response (i.e., they indicated a powerful group as powerful). In the gender block, we also recorded pupil size on all types of trials (male/female names categorized as powerful at the top or bottom).

Procedure

After providing consent, participants sat in front of the computer screen (approximately 67cm away) and their eye position in the eye-tracker was calibrated. We used E-Prime 2.0 software to present all stimuli, record responses, and measure pupillary activity registered by the eye tracker (Psychology Software Tools, Pittsburgh, PA). After the calibration was complete, participants were instructed to quickly categorize the powerful person on the screen for both the Schubert and gender trials (see Figure 5). The instructions were the same as for the previous studies (i.e., they included a definition of a socially powerful person). Participants were tested in a dark room. The experiment lasted approximately 45 minutes.

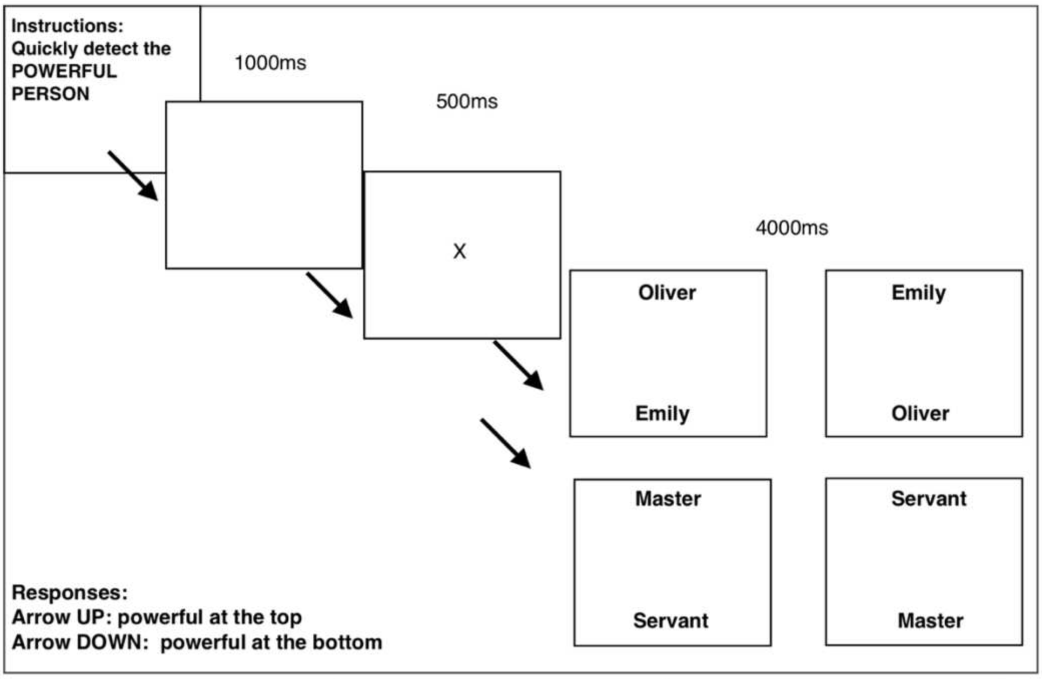


Figure 5. Spatial task procedure.

Results

Pupil size pre-processing.

First, all of the gaze data that were missing due to blinks or incorrect recordings (as computed and indicated by E-Prime software) were coded as missing values. These constituted 22% of the data in the Schubert block and 18% in the gender block (Funke et al. 2016 estimated that Tobii eye trackers output 78% of useable data). Second, we combined the pupil data from the left and right eye into one pupil size score that represented the mean pupil size for both eyes. Following the procedure used by Proulx et al. (2017), we filtered the pupil size signal with a repeated median regression filter[20] (robfilter package by Fried, Schettlinger, & Borowski, 2014) using statistics software R (R Core Team, 2016) to smooth implausible pupil recordings. Such invalid measurements might be due to participants' head movements during the task, as no chin rest was used.

Once the pupil size signal was smoothed, the missing signal that was not likely associated with eye blinks (as its duration was too short, i.e., fewer than four data entries, which equal to 33 ms) was corrected by a linear interpolation.[21]

Subsequently, we removed artifacts associated with blinks. Blinks are recorded as missing data by the eye-tracker, however, they also lead to incorrect measurements around 100 ms before they occur (as participants' eyelids close, the eye tracker fails to record the entire pupil). Also, incorrect measurements are likely to be recorded 200 ms post-blink. This is caused by opening of eyelids as well as an initial constriction of the pupil that occurs immediately after each blink (Lin et al., 2018). Therefore, all pupil recordings 100 ms pre-blink and 200 ms post-blink were removed. The missing signal due to blinks was filled in by a linear interpolation. The missing signal recorded for more than 500 ms was not interpolated, as blinks do not last longer than this period. Overall, after pre-processing, the estimated data loss was reduced to 11% in the Schubert block and 9% in the gender block.

Finally, to correct for baseline differences in pupil size between participants, we calculated the mean pupil diameter during 100 ms of the fixation period that occurred just before the target trial period presentation, following Mathôt, Fabius, van Heusden, and van der Stigchel's (2018) suggestions (the target trial period refers to the trial event when pairs of powerful-powerless groups and female-male names appeared on the screen). This was done for each participant and each trial. Then, we subtracted that mean score from each

[20] In general, median filters inspect time-series signal entry by entry within a pre-determined window (a window indicates a pattern of adjacent values) and replace the data points within that window with a median of entries adjacent to the window.

[21] The linear interpolation replaces missing signal with the mean of neighbouring valid signal measurements.

subsequent pupil size measurement during the target trial period.[22] After the baseline corrections, all pupil size values that were closer to 0 represented a normal pupil size for participants before each trial began.

Primary Analyses

To begin, we computed the mean pupil size for each participant, trial, trial type (men-top versus women-top), and gender that was indicated by participants to be powerful (gender choice: male versus female names). Then, we calculated the mean response time (M = 1415, SD = 666) across all participants and trials to obtain a reference in order to analyze participants' pupil size in the pre-response and post-response periods.

Pre-response period. We first tested whether the simple presentation of women-top trials versus men-top trials (regardless of participants' subsequent gender choice) would involve increased pupil size in the context of the task instructions (i.e., categorizing the powerful person). If pupil size increases are due to unexpected stimuli presentation, such increased dilation should be present in the pre-response period regardless of participants' choice. Figure 6 suggests that the biggest difference in pupil size was observed in the case of 750 ms time-stamp. As hypothesized, participants' pupil dilation was significantly higher for women-top trials (M = -.14mm, SE = .05, 95% CI[-.24, -.04]) compared to men-top trials (M = -.19mm, SE = .04, 95% CI[-.28, -.08]), $F(1,590) = 4.81$, $p < .026$, $dz = .57$ (95% CI[.03, 1.09]; n = 16), signaling more surprise on women-top trials. Also, as expected, this effect was independent of participants' later response within the trial, which is demonstrated by the lack of significant interaction between gender choice and trial type, $F(1,593) = 1.35$, $p = .176$. As demonstrated by Figure 6, and confirmed by additional analyses, the differences in pupil

[22] As the literature does not report one valid way of pupil pre-processing techniques (e.g., Cavanagh, Wiecki, Kochar, & Frank, 2014; Lin et al., 2018; Mathôt et al., 2018), we combined different smoothing and filtering techniques described in the literature that best fit our data in terms of signal quality, signal loss, eye-tracker speed, and trial length. We reported only one technique, however, using different window widths in median filters or not applying blinks corrections led to similar trends in the data analyses.

sizes across the trial types within the whole pre-response period (500-1415 ms) as well as individual time-stamps of 500 ms and 1000 ms, were not significant, *ps* > .10. Such findings indicate that there was initial increase in pupil size at 750 ms in the women-top trials, suggesting that presenting women at the top evoked cognitive conflict or surprise. This is also supported by the results showing that such an increase in pupil size was not associated with the response that participants made later within the trial. After 750 ms, pupil size within both conditions increased to the same extent most likely as a function of general cognitive load associated with decision-making.

Furthermore, as indicated by Figure 7a, it seems that participants demonstrated increased pupil size when they were presented with men-top trials and subsequently selected the female name as powerful. The differences were most pronounced at 500 ms and 750 ms and are still visible in the post-response period. The same difference, however, is not demonstrated by Figure 7b, suggesting no differences in pupil size between choices of women (at the top) or men (at the bottom) on women-top trials. To test whether these differences were meaningful in the context of subsequent decision-making, we conducted additional analyses. Post-hoc comparisons indicated that the difference in pupil size between choices of men when they appeared at the top and women when they appeared at the bottom at 500 ms was marginally significant, $t(596) = 1.80$, $p = .072$, $dz = .45$ (95% CI[-.06, .96]; $n = 16$), such that pupil dilation at 500 ms was increased when women were subsequently chosen at the bottom ($M = -.12$mm, $SE = .05$, 95% CI[-.28, -.06]), in contrast to choosing men at the top ($M = -.25$mm, $SE = .05$, 95% CI[-.32, -.11]). This is consistent with previous literature suggesting that when people decide to control their initial tendencies, like in our task categorizing women as powerful at the bottom (when a choice of men or top is available), they need to inhibit the powerful-top association, and/or the stereotype associated with men as powerful first (van der Wel & van Steenbergen, 2018). This in turn involves pupil enlargement. The same, however, was not true at 750 ms, as the interaction between gender choice and trial type at that time-stamp was not significant, $F(1, 590) = 1.83$, $p = .176$; $n = 16$. The same differences in pupil size later on within the trial were not significant at individual time stamps, *ps* > .19. Finally, comparisons between choices of men at the bottom versus women at the top on women-top trials were not significant, *ps* > .33.

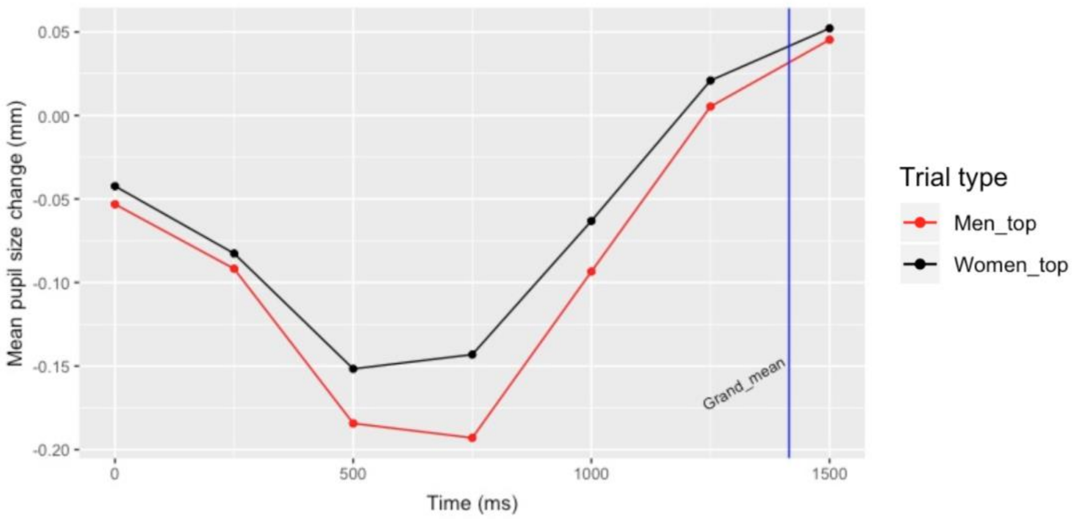
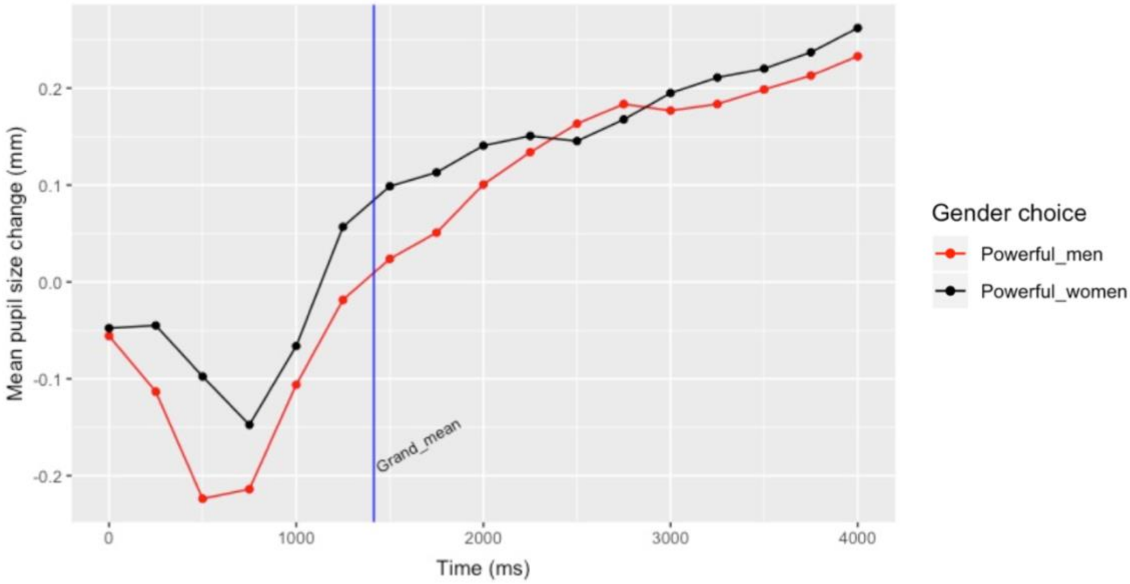


Figure 6. Mean pupil size change as a function of presenting men-top and women-top trials in the pre-response period of the gender block. The vertical line indicates the mean response time across all participants and conditions (M = 1415 ms, SD = 666). Participants were instructed to look for the powerful person.

(a) Men-top trials



(b) Women-top trials

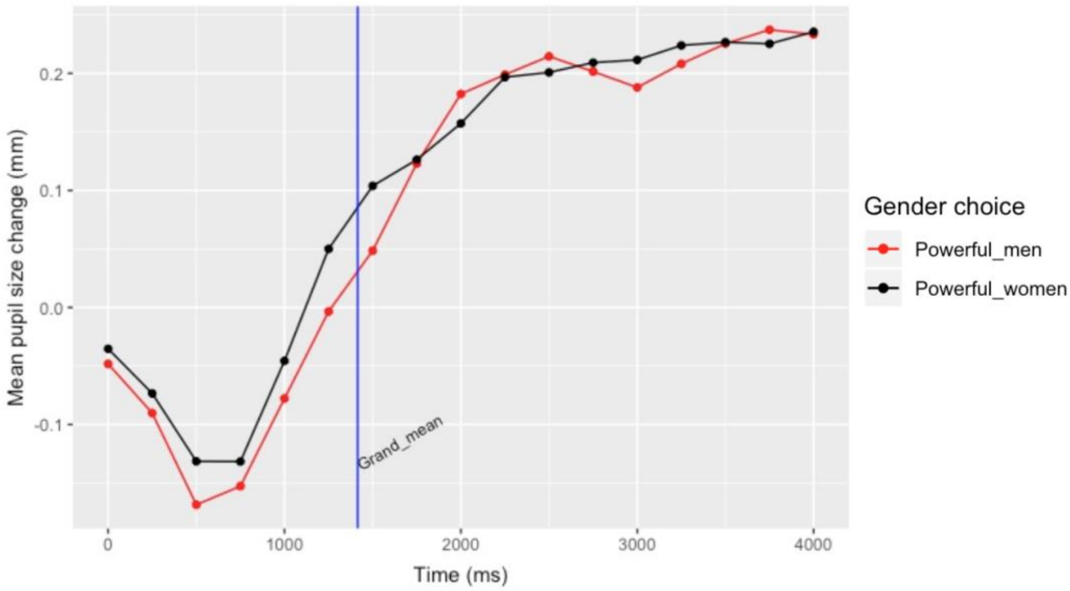


Figure 7. Mean pupil size change as a function of presenting (a) men-top trials and choices of male and female names on that trials (i.e., men at the top; women at the bottom; both chosen as powerful) and the same choices are presented for (b) women-top trials over 4000 ms. The vertical line indicates the mean response time across all participants and conditions (M = 1415 ms, SD = 666).

Entire-trial period analysis. Building upon the results above, we also analyzed participants' pupil dilation across the entire trial period (500-4000 ms). We found that the main effects of gender choice, $F(1,13) = .76$, $p = .409$, $dz = .21$ (95% CI[-.28, .70], $n = 16$), and trial type, $F(1,1168) = .97$, $p = .320$, $dz = .25$ (95% CI[-.25, .74], $n = 16$), were not significant. The main effect of response period was significant, $F(1,15) = 37.01$, $p < .001$, $dz = 1.52$, (95% CI[.80, 2.23]; $n = 16$) with increased pupil dilation in the post-response period, ($M = .17$mm, $SE = .04$, 95% CI[.07, .26]), than in the pre-response period, ($M = -.09$mm, $SE = .04$, 95% CI[-.17, -.01]). As expected, the two-way interaction between trial type and gender choice was significant, $F(1,15) = 8.23$, $p < .003$; $n = 16$. On trials where participants selected men as powerful, their pupil size was significantly increased when the male name appeared at the bottom ($M = .05$mm, $SE = .03$, 95% CI[-.01, .13]) as compared to the top ($M = .01$mm, $SE = .03$, 95% CI[-.06, .07]), $t(1190) = 3.19$, $p < .008$, $dz = .80$ (95% CI[.23, 1.36]; $n = 16$; see Figure 8). However, the same was not true on trials where participants indicated women as powerful, with no difference in pupil size when the female name appeared at the bottom of the screen ($M = .05$mm, $SE = .03$, 95% CI[-.03, .17]) compared to the top ($M = .06$mm, $SE = .05$, 95% CI[-.06, .14]), $t(1199) = 1.22$, $p = .615$, $dz = .30$ (95% CI[-.20, .81]; $n = 16$; see Figure 9). All remaining interactions were not significant (*ps* > .61).

Further, as indicated by Figure 10, participants' pupil dilation was increased when participants judged women as powerful on women-top trials ($M = .060$mm, $SE = .05$, 95% CI[-.03, .17]) as opposed to men as powerful on men-top trials, ($M = .006$mm, $SE = .04$, 95% CI[-.06, .13]), $t(1205) = 1.96$, $p < .049$, $dz = .49$ (95% CI[-.03, 1.00]; $n = 16$). Consistent with Figure 11, there was no significant difference between pupil dilation when participants indicated bottom positions on men-top trials (i.e., when women as powerful were chosen at the bottom; $M = .05$mm, $SE = .05$, 95% CI[-.01, .19]), and women-top trials (i.e., when men as powerful were chosen at the bottom; $M = .05$mm, $SE = .05$, 95% CI[-.03, .17]), $t(1206) = -$

.17, *p* = .860, *dz* =.04 (95% CI[-.44, .53]; *n* = 16). Overall, these results imply that across the whole trial period, participants exhibited increased pupil size, and hence higher cognitive effort, when they selected a counter-stereotypic stimulus (i.e., women as powerful) at the top as opposed to a stereotypic stimulus (i.e., men as powerful) at the top. Further, there were no differences in pupil size when selecting counter-stereotypic stimuli and stereotypic stimuli at the bottom, suggesting that cognitive load associated with those choices was comparable.

Post-response period. Finally, because participants could make any type of choice (men and women as powerful when presented at the top or bottom), we tested whether the type of choice they made was associated with differential response-locked cognitive effort. We thus conducted the same analysis as presented above, however, this time we analyzed participants' pupil size only after they made their response (i.e., in the period > 1415 ms). Similar to the analysis of the whole trial period, we found that the main effects of gender choice and trial type were not significant, *ps* > .72. However, as expected, the gender choice by trial type interaction was significant, $F(1, 595) = 5.23$, $p < .021$; $n = 16$. The post-hoc comparisons replicated the results of the previous analysis, indicating increased pupil dilation when men were chosen as powerful when they appeared at the bottom of the screen (*M* = .18mm, *SE* = .04, 95% CI[.11, .27]) as compared to the top (*M* = .14mm, *SE* = .04, 95% CI[.05, .21]), $t(600) = 2.35$, $p < .019$, *dz* = .59 (95% CI[.06, 1.12]; *n* = 16) The same was not true for women chosen when they appeared at the bottom of the screen (*M* = .16mm, *SE* = .05, 95% CI[.08, .31]) versus the top (*M* = .18mm, *SE* = .05, 95% CI[.05, .27]), $t(600) = 1.12$, $p = .260$, *dz* = .28 (95% CI[-.22, .78]; *n* = 16). Further analyses did not replicate the findings from the entire-period results, indicating that participants' pupil size was not significantly different when participants indicated top or bottom stimuli across men-top and women-top trials, *ps* > .25. The reported results within both blocks were not associated with participants' gender or counterbalancing of blocks, *ps* > .20.

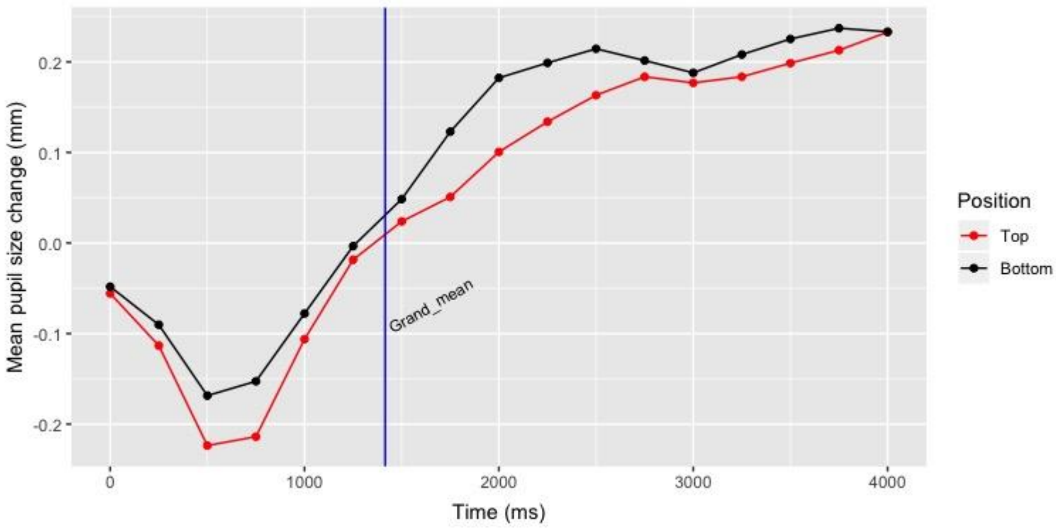


Figure 8. Mean pupil size change as a function of choosing men as powerful when their names appeared the top and bottom over the trial period (4000 ms). The vertical line indicates the mean response time across all participants and conditions (M = 1415 ms, SD = 666). Participants were instructed to look for the powerful person.

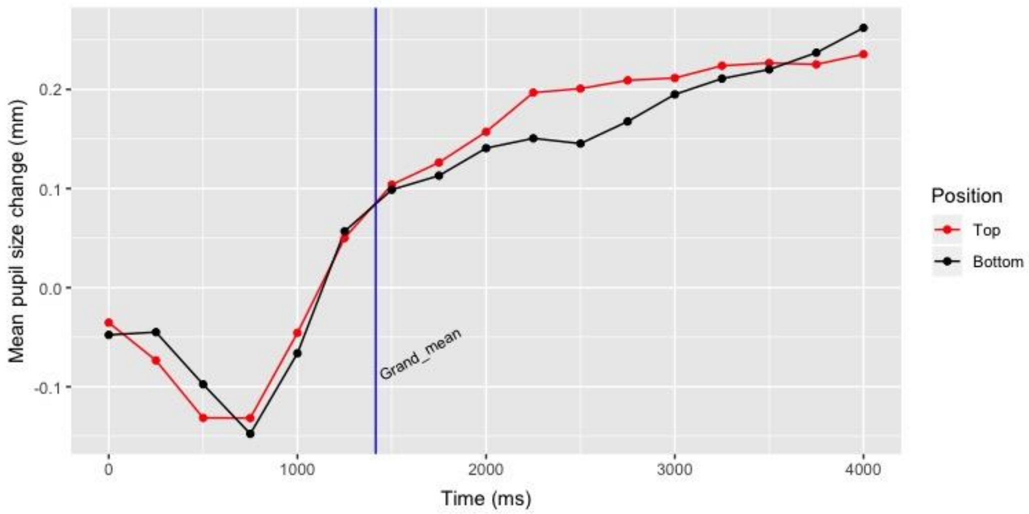


Figure 9. Mean pupil size change as a function of choosing women as powerful when their names appeared at the top and bottom over the trial period (4000 ms). The vertical line indicates the grand mean of response time.

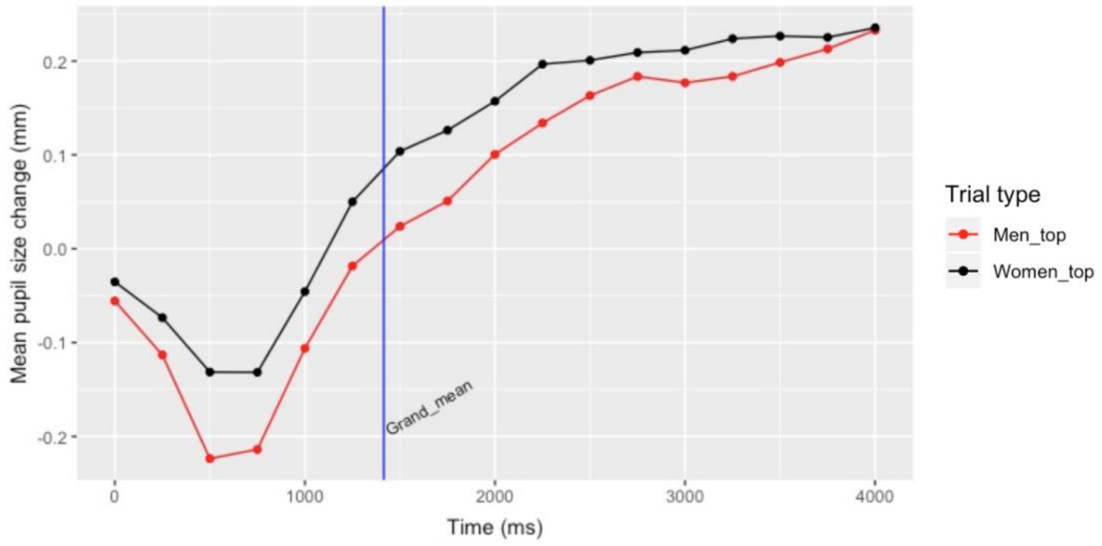


Figure 10. Mean pupil size change as a function of choices of men and women as powerful at the **top** (*Men_top* indicates men-top trials and the choice of men, whilst *Women_top* indicates women-top trials and the choice of women).

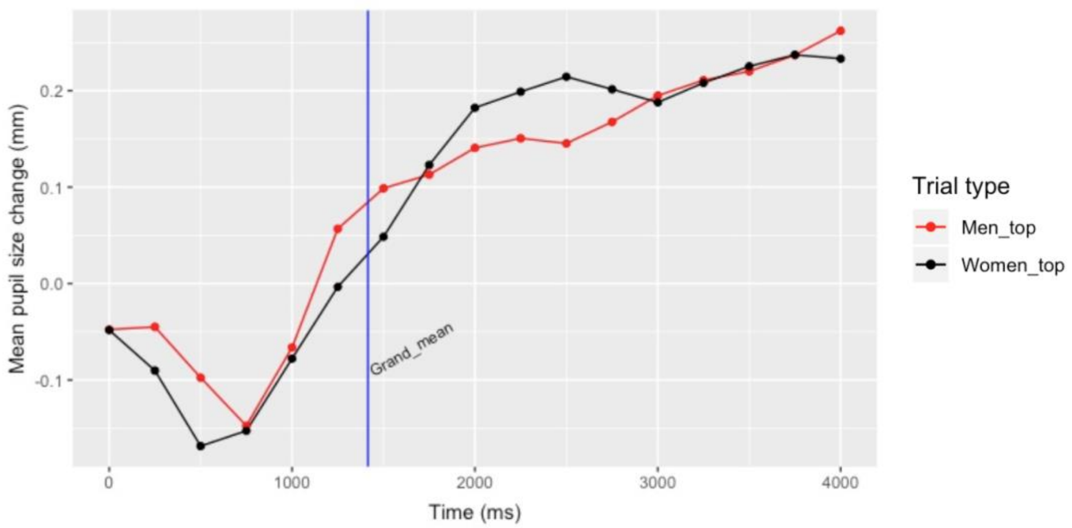


Figure 11. Mean pupil size change as a function choices of men and women as powerful when their names were selected at the **bottom** (*Men_top* indicates choices of women at the bottom, whilst *Women_top* indicates men chosen at the bottom).

Study 7: Discussion

We designed Study 7 to assess whether the stereotype-metaphoric interaction associated with gender categorizations could be detected at the pre-conscious level, independent of behavioral responses. We used a physiological recording of pupil size to test this idea. First, in order to assess whether pupillometry was appropriate to test our hypotheses, we conducted a replication of Schubert's study (2005). We asked participants to complete a spatial task, whereby they indicated which person among two vertically presented ones (e.g., *master* or *servant*) was powerful (see supplementary analyses for results of this block). Second, we conducted a replication of our own behavioral paradigm, in which participants' task was to decide which person was powerful among two vertically presented male and female names. In both replications, we recorded participants' pupil size while they were completing the tasks.

Schubert block

Consistent with our hypotheses, we found that across the whole trial period, participants exhibited increased pupil size on trials where they were presented with powerful groups at the bottom as opposed to the top. The same pattern was noted at the pre-response 750 ms time-stamp, however, in this case the increase in pupil size on the incongruent trials as opposed to the congruent ones was marginal. These findings suggest that participants experienced higher cognitive conflict when they processed powerful groups at the bottom. In the post-response period, we found that this difference was significant immediately after participants made their responses (i.e., in the period from 1379 ms to 2000 ms). This effect most likely reflects response-locked cognitive effort (van der Wel & van Steenbergen, 2018).

Gender block

Consistent with our hypotheses, we found that participants' pupil size was associated with an interaction between gender choice and trial type. Within the whole response period, we found increased pupil size on trials when men were selected as powerful at the bottom as opposed to the top - supporting the findings of our previous studies and the Schubert block. The same results were obtained when the post-response period was analyzed separately.

Our analyses of the pre-response period suggest that participants experienced an expectancy violation when they were presented with trials on which female names appeared at the top and male names at the bottom. This is because participants demonstrated increased pupil size when presented with trials where female names appeared at the top as opposed to male names in the same location and such pupil size dilation was *independent* of participants' response later on within the trial. The results are in line with the findings obtained in the Schubert block showing that participants also experienced increased pupil size in the pre-response period on trials presenting powerful groups at the bottom as opposed to the top. Interestingly, such differences were noticed at 750 ms in both blocks (although the difference was marginal in Schubert's block). Because the differences were independent of participants' later decisions about men or women as powerful in the gender block, it appears that higher pupillary dilation associated with observing female names presented at the top (in the context of looking for a powerful person) was related to increased cognitive conflict. Such cognitive conflict seemed to be responsible for the increased pupil size in the early period of the trial and this supports previous research (Lin et al., 2018; Preuschoff et al., 2011; Proulx et al., 2017; Sleegers et al., 2015).

When analyzing the data from the entire-trial period, participants exhibited marginally higher pupil dilation when they selected female names when they appeared at the top as opposed to male names at the top. Again, this difference supports our hypotheses and the behavioral findings from Studies 1–6, suggesting higher cognitive effort when processing stereotype-inconsistent stimuli at the top (i.e., powerful-women). These findings are novel and important, as it appears that choices of stereotype-consistent (powerful-men) and –

inconsistent (powerful-women) targets per se do not involve expectancy violation and hence higher pupil dilation. Instead, mental conflict observed in our study was specifically associated with spatial location interacting with stereotypic thinking.

Furthermore, it is important to note that the findings of the difference observed in pupil size between choices of women as powerful at the top as opposed to men as powerful at the top were not found when post-response period was analyzed separately. This is surprising because processing of inconsistent information that requires inhibition of the interfering distractor should be associated with increased pupil size (Aston-Jones & Cohen, 2005, Laeng et al., 2011). It is possible that the general increase in pupil size in the entire trial within our study was primarily driven by the initial and response-preparatory cognitive conflict of perceiving women at the top while tasked with selecting the powerful person. These results could be also attributed to the fact that participants could not compare both men and women at the top within the same trial. Therefore, processing of women at the top would be mainly associated with distinguishing them from men at the bottom.

In fact, our additional analyses of differences between choices of male and female names in different trial types indicated that participants exhibited marginally higher cognitive load on trials when they picked female names as powerful at the top as opposed to male names at the bottom. Interestingly, the increased pupil size to the presentation of female names at the top at 500 ms, but not 750 ms, was marginally associated with later categorizations of those women as powerful. Further, it can be speculated that the increased pupil size at this early stage, which was most likely associated with aversive arousal to an inconsistent stimulus (women at the top), stimulated participants to engage in compensatory behaviors and hence counter-stereotypic choice. Such findings are in line with suggestions by Proulx, Inzlicht, and Harmon-Jones (2012), who argue that aversive arousal/cognitive conflict associated with unexpected stimuli is compensated by either accommodation or assimilation of counter-attitudinal or stereotypic information. However, these results are marginal, and they do not imply any causal link between increased pupil size and later categorizations of female names as powerful at the top. Therefore, these suggestions should be treated with caution.

In sum, we further demonstrated that people hold metaphorical expectations about social power and these expectations interact with stereotype-consistent beliefs. These novel findings fully complement our previous behavioral studies and present extra evidence for the interaction between stereotypes and metaphors at the level of pre-conscious expectations.

General Discussion

In the present manuscript, we show initial evidence for the simultaneous employment of two interacting mental strategies in social judgement. We extend the previous literature by demonstrating that people use stereotypes in combination with other mental shortcuts to make social categorizations. Specifically, stereotypic associations between gender and power can be conceptually combined with space-power metaphors, such that stereotypically powerful individuals (i.e., men) are represented in the metaphorically congruent spatial position, that is, at the top. We further supported these findings by demonstrating that stereotypes interact with metaphors even at the level of pre-conscious processing and this interaction creates combined stereotype-metaphoric expectations concerning characteristics of gender – that is, men as powerful should be located at the top rather than bottom.

On the basis of our findings, we argue that while making social categorizations, people activate one coherent mental representation that comprises of two associated mental shortcuts. In other words, a stereotypical judgment about men as powerful seems to be accompanied by a spatial simulation in upper vertical positions. Stereotypes and metaphors aid cognitive processing of information, such that stereotypes reduce the complexity of social information (Fiske & Taylor, 2010), whilst metaphors illustrate abstract concepts by concrete ones to facilitate their understating (Lakoff & Johnson, 1999). Under such circumstances, it is likely that the stereotype-metaphoric interaction facilitates people's social judgments. Overall, we wish to emphasize that our findings provide evidence for the interplay between metaphors and stereotypes. Further research is needed to empirically examine whether such stereotype-metaphor blend can in fact improve the *efficiency* of cognitive processing.

We propose that the conceptual blending perspective (Fauconnier & Turner, 1998) is useful in illustrating the mechanisms that potentially underlie the association among power, verticality, and gender. This is because conceptual blending refers to a process of representing multiple associated concepts in a coherent mental representation, as opposed to representing each concept separately. In our studies, participants exhibited a processing advantage whenever they selected a man at the top when considering him as powerful. That is, stereotypic links between power and gender blend men with power. Because metaphorically power is associated with verticality (Schubert, 2005), men-powerful links become associated with the metaphoric characteristics of power, that is, verticality. This process creates a unified mental model that men who are powerful are at the top. Consistent with this reasoning, Maass, Suitner, Favaretto, and Cignacchi (2009) demonstrated that people in Western cultures (with left-to-right writing habits) spontaneously draw male targets to the left of female targets when considering men as agentic. This illustrates conceptual blending between men and agency. As agency is associated with *left*, people have a coherent mental model of agentic men on the left. If such blending among concepts did not occur, we would predict that one mental strategy would take priority over the other. For example, if people used only gender stereotypes (e.g., men-powerful; men-agentic), we would expect them to be quicker at indicating powerful men at the top and at the bottom of the screen (or draw agentic men sometimes on the left or and sometimes on the right). At the same time, if people only used metaphors of powerful-top/agentic-left, we would expect them to quickly detect (or spontaneously draw) all powerful and agentic targets at the top and left, respectively, independent of gender. However, empirical evidence from our studies and Maass et al. (2009) demonstrates that linked concepts are in fact represented in a unified mental model by the means of conceptual blending.

Our follow-up study (see the Online Supplemental Materials) further supports the perspective of conceptual blending in illustrating that the gender-power-verticality associations are represented in a unified mental model. In this study, we demonstrated that in the absence of the power context, participants were equally fast at categorizing gendered names as male and female when the names were presented at the top and bottom of the screen. Therefore, verticality is not directly associated with gender. This further indicates that it is unlikely that our findings were determined by other types of gender associations, such as physical height. As men are on average taller than women, it was possible that men were directly associated with upper location. Because we did not detect such links in the follow-up study, it is unlikely that stepwise or additive associations between men-top and men-power were responsible for our results. This is likely in the context of Montepare's (1995) results who found that perceptions of the relationship between height and dominance/strength were not moderated by the target's gender.

In sum, the behavioral studies demonstrated that participants were significantly quicker at detecting men as powerful when their names appeared at the top as opposed to women as powerful at the top. To further support this finding, we additionally found that faster reaction times predicted a higher probability of selecting men, only in the powerful condition and on men-top trials. As well, this effect was associated specifically with vertical location, as men were categorized as powerful significantly faster when they appeared at the top than bottom. The same pattern of results was not true for female names. They were categorized as powerful equally fast at top and bottom.

The results of Study 7 further confirmed our predictions that stereotypic expectations interact with linguistic metaphors and this effect was independent of behavioral responses. We tested this by assessing the stereotype-metaphoric expectancy violation at a pre-conscious level. First, we demonstrated that pupillometry can indeed assess metaphor-consistent expectations, by replicating Schubert's (2005) results. We found that participants exhibited increased pupil size when they categorized powerful groups at the bottom of the screen as opposed to the top at the pre- and post-response periods demonstrating an initial expectancy violation and a later response-locked cognitive effort. Subsequently, we found that the same pattern was observed when participants were presented with a female name at the top of the screen compared to a male name at the top, when they were asked to categorize the powerful person. This effect was independent of participants' behavioral categorizations later within the trial demonstrating the violation of stereotype-metaphoric expectations of men at the top.

We also observed significantly increased pupil size on trials when participants categorized men as powerful at the bottom versus the top, which reflected increased response-locked cognitive load. Our findings are in line with the pupillometry literature, indicating that early pupillary dilations are associated with expectancy-violation responses related to perceiving social, attentional, and semantic incongruencies (Lin et al., 2018; Preuschoff et al., 2011; Proulx et al., 2017; Sleegers et al., 2015; van Steenbergen & Band, 2013). In addition, our observation of response-locked pupil dilations supports previous multiple studies demonstrating a positive correlation between cognitive effort and pupil size (see van der Wel & van Steenbergen, 2018). These novel results build upon previous research by further indicating that conceptual blending enables the metaphoric space-power interaction with gender stereotyping creating pre-conscious expectations of men at the top. Overall, our reaction-time data converge with pupillometry results providing a strong support for the stereotype-metaphoric interaction in human social categorizations.

Our research further contributes to the literature by building on the notion that representations of space play an important role as cognitive scaffolds and working metaphors in social perception and reasoning. We suggest that links between space and power as well as gender-power stereotypes interact by means of conceptual blending. If conceptual blending did not occur in our studies, we would expect that whenever high power was attributed to either female or male targets, the simulations would occur, such that both genders would be simulated at the top. In the context of attributing low power, again, both female and male targets should be simulated at the bottom. That is, in this case the power-space metaphor would be independent of stereotypic associations. This perspective would be in line with Spatola et al. (2018), who demonstrated that although multiple metaphoric associations can be activated simultaneously, they do not interact with each other in categorization tasks. In the current paper, we extend these perspectives in two important ways. First, we show that metaphors can indeed interact with real-life social judgements about specific individuals that are semantically relevant to the metaphor. Second, although metaphors do not interact with other metaphors (Spatola et al., 2018), they interact with stereotypes. We demonstrate that at *both* pre-conscious and behavioral levels, people hold expectations based on the combined space-power metaphor and the gender-power stereotype. Further, it is important to note that the spatial simulations of men as powerful in the upper vertical location, as demonstrated in our studies, did not require deliberate conscious processing. As shown by Study 7, the violation of pre-conscious expectations, whereby male targets presented in the power context and located at the bottom of a display, were responsible for the early-pre-response increased pupillary dilations.

The implication of our findings is that when attributing high power to social categories, it is easier to process male names when they are presented at the top as opposed to female names at the top. It is also easier to process male names presented at the top as opposed to the bottom. Such spatial links might in turn be associated with an increased readiness to activate and apply gender stereotypes in situations where power context is relevant. Thus, it is possible that the men-top-power association act as a generalized, distinct, and implicit mental shortcut facilitating stereotype-consistent decisions, for example, when selecting and creating a rank order of job applicants for high-status positions, evaluating employees' performance, or considering them for promotions.

In terms of broader implications, it is possible that people vary in the extent to which they are able to use metaphors and stereotypes simultaneously. Potentially, under conditions of high ambiguity, people with a high ability to combine mental shortcuts might be more confident in making decisions and be more satisfied with their choices than people who are less able or motivated to use two strategies. This is because such shortcuts represent multiple sources of information that could be used to inform decisions. Overall, using two shortcuts as opposed to one might facilitate difficult decision making.

The bottom-powerless metaphor does not interact with stereotypes

In the behavioral studies, no effects emerged in the case of selecting male and female names in the powerless condition. This is not in line with Schubert's (2005) findings, as he found spatial simulations of powerless individuals at the bottom. That said, Schubert found that such effects were weaker relative to powerful individuals (see also Meier & Robinson, 2004; von Hecker et al. 2016). There are numerous plausible explanations for the lack of spatial simulations of women as powerless. First, substantial research findings provide strong support for the notion that people better remember and pay more attention to high-status compared to low-status individuals (see Mattan, Kubota, & Cloutier, 2017, for a review). Social cues associated with high power and social rank orders were found to be processed automatically at the pre-conscious level in studies where power was not relevant to the task conditions (Zink et al., 2008). Such findings demonstrate that people exhibit automatic readiness to identify social power-related cues (Chiao et al., 2009; Stewart et al., 2012). In the context of our findings, it is possible that participants who were asked to categorize powerless individuals paid attention to powerful targets first.

Second, research has demonstrated that spatial effects are usually more pronounced for linguistically unmarked concepts (e.g., powerful) than for marked ones (e.g., powerless). Unmarked concepts refer to concepts that describe a dimension entirely, such that asking how *powerful* someone is might reflect high versus low levels of power. Conversely, asking how *powerless* someone is indicates that the person can be only powerless, but not powerful. In terms of space-power metaphors, we usually talk about *power is up*, the *top dog* rather than *powerless is bottom* or the *bottom dog*. This is in line with the idea that unmarked concepts (i.e., powerful) are processed faster than marked concepts (i.e., powerless) due to familiarity, evaluative implications, and higher linguistic frequency (Hamilton & Deese, 1971; Meier & Robinson, 2004; Meier, Hauser, Robinson, Friesen, & Schjeldahl, 2007; Proctor & Cho, 2006; Schubert, 2005; von Hecker et al., 2016). Taken together, the gender stereotypes interaction with space-power metaphors is based upon the heightened attentional bias towards powerful targets as well as the linguistic markedness of the word *powerful.*

Spatial simulations and implicit attitudes

In terms of the attitudes towards high-status men/women, participants tended to associate men with high status and these associations were not related to spatial simulations of either men or women. Turning to the rationality/emotionality and gender IAT, consistent with our hypothesis participants held associations between men and rationality, but not between women and rationality. Such associations were not related to spatial simulations of either men or women. The same was true for top-rationality associations (inconsistent with Cian et al., 2015). Participants did not exhibit response latency facilitation to top-rationality items as opposed to top-emotionality items. Oswald, Mitchell, Blanton, Jaccard, and Tetlock

(2013), in their meta-analysis indicated that IATs are weakly associated with behaviors, judgments, and explicit attitudes. In addition, Greenwald, Poehlman, Uhlmann, and Banaji (2009) showed that IATs had poor predictive validity in the domain of gender biases. However, implicit attitudes highly predict behaviors in cases where there is a high implicit-explicit attitudes correlation (Jost, 2019; Kurdi et al., 2018). Research has also demonstrated that IATs partially reflect environmental influences (i.e., biases at a regional/country level; Payne, Vuletich, & Lundberg, 2017). Hence, it is possible that the IATs used in our studies did not necessarily tap into *participants'* underlying stereotypic associations. The role of individual differences in the stereotype-metaphoric blend should be therefore further investigated by employing explicit gender bias scales.

Women are not spatially processed

We detected spatial simulations only in the context of thinking about men as powerful, but not women as powerless. Indeed, previous literature indicates that spatial simulations are not consistently detected across all contexts (e.g., Borghi, Glenberg, & Kaschak, 2004; Bub, Masson, & Cree, 2008; Lebois et al., 2015). Our findings extend this perspective in a novel and important way by suggesting that an application of the power concept to social groups, and especially stereotypic expectation about those groups (men-powerful), enables spatial processing. This is based on the extent to which power is attributed to gender in the presence of stereotype-fit between *men* and *power*. That is, the men-power link tends to be salient and guides participants' responses in the context of the choice task (see also Rudman & Kilianski, 2000). Indeed, DeWall and Maner (2008) demonstrated that people pay more attention to high-status men rather than women. This likely explains the lack of effects when people were asked to look for powerless individuals. Even though the stereotype-fit was present when participants selected women as powerless, such choices did not involve spatial processing, as participants did not simulate women as powerless at the bottom – even though their response pattern indicated a slight advantage (however, not significant one) at choosing women as powerless at the bottom as compared to men as powerless at bottom.

In relation to our previous arguments, we would like to note that our spatial reaction-time test was a forced-choice task in which participants were required to provide a response, even in situations when they might not have found any of the two individuals particularly powerful/powerless. Hence, random fluctuations in the salience of the stereotype might have taken place during the experiment across trials. The random fluctuations argument finds support in a recent review article on temporal instability of implicit biases. As hinted above, Payne et al. (2017) argue that situational factors rather than stable individual attitudes are more likely to determine the accessibility of stereotype-consistent content in people's minds that lead to stereotype-consistent responses. Such reasoning seems logical in the context of research showing low test-retest correlations in individual levels of implicit prejudice in longitudinal studies (see Gawronski, Morrison, Phills, & Galdi, 2017). Also, social categories in cognition are typically abstract and may subsume variable sets of exemplars. Therefore, the context of thinking might affect accessibility of concepts and exemplars, and might be more important in determining which associations will be retrieved when making a particular judgment (Medin, 1989). Under such conditions, the context of our task possibly may have stimulated participants, in some trials, to activate the power concept in a counter-stereotypic situation (i.e., considering women as powerful), such that the spatial features of power, in such trials, may then have suggested a top-position for the female. Overall, we speculate that the mapping of abstract concepts onto the concrete dimension of space is particularly likely under situational factors that engender a high (stereotypical) consistency between the concepts involved.

Future directions

It is worth noting that all of our studies are based mainly on samples of female participants. As described earlier, the samples were representative of their participant panels. Previous research has indicated that women have equally negative implicit attitudes towards powerful women as men, but they are less explicitly prejudiced against women as powerful than men (Rudman & Kilianski, 2000; Rudman et al., 2012; Rudman, & Phelan, 2010). We strongly believe that spatial simulation in cognitive blending is a general phenomenon across gender (and it appears that it is independent of people's implicit attitudes, as found in our studies; see also Roets, van Hiel, & Dhont, 2012). In addition, Hanel, Maio, and Manstead (2019) demonstrated that there was 97% of similarity between men and women in terms of human values as well as social and political attitudes across many countries. In our context, it would be valuable to experimentally test whether men and women exhibit differential spatial simulations of men as powerful and women at the top directly. That said, considering available literature and to the best of our knowledge, the effects we detected are not limited to any particular population.

This is further supported by the obtained effect sizes (Cohen's dz = .33 - .80) that indicate solid evidence for the spatial effects we detected across the behavioral and physiological studies. We found especially large effect sizes in the pupillometry experiment (i.e., Cohen's dz = .57 - 80). Such effects are larger than typical effect sizes detected in social psychology research (i.e., Cohen's d = .45; Richard, Bond Jr & Stokes-Zoota, 2003). In the behavioral experiments, the effects were less pronounced (Cohen's dz = .33 - .35) than in the pupillometry study. This might suggest that applying reaction-time methodology is associated with an ability to detect more subtle spatial effects than the physiological technique. Overall, combining these two methodologies allowed us to provide robust evidence for the investigated effects.

Finally, we wish to note that the conclusions drawn from six initial studies are based on an integrative analysis. As not all individual studies (see Appendix F) achieved standard significance levels, it is important to acknowledge that while testing the same hypothesis across several studies, it is unlikely that each of the studies would support that prediction, and that an integrative analysis provides more reliable results (see e.g., Lakens & Etz, 2017).

Conclusion

In this paper, we demonstrate that people have an ability to use more than one mental strategy while making social judgements. Specifically, people simultaneously employ two mental shortcuts given that these shortcuts can acquire each other's properties by means of conceptual blending. We illustrated this process in our research, where stereotypic associations acquired features of linguistic metaphors in a gender-categorization task. We argue that blending the concept of power with social concepts that have no immediate power implications (i.e., gender should be in principle power-neutral), but evoke stereotypic expectations, is sufficient to activate spatial processing associated with simulations of power on the vertical dimension.

References

Aston-Jones, G. & Cohen, D. J. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance, *Annual Review of Neuroscience, 28*, 403-450.

Barhum, L. (2018, March 6). Average height for women worldwide. Retrieved June 16, 2019, from https://www.medicalnewstoday.com/articles/321132.php

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition, 32*, 863–873.

Bub, D. N., Masson, M.E.J., & Cree, G.S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106,* 27-58.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General, 138,* 352-367.

Cavanagh, J.F., Wiecki, T.V., Kochar, A., & Frank, M.J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General, 143*, 1476–1488.

Chiao, J. Y., Harada, T., Oby, E. R., Li, Z., Parrish, T., & Bridge, D. J. (2009). Neural representations of social status hierarchy in human inferior parietal cortex. *Neuropsychologia*, *47*, 354-363.

Cian, L., Krishna, A., & Schwarz, N. (2015). Positioning rationality and emotion: Rationality is up and emotion is down. *Journal of Consumer Research, 42,* 632-651.

Clark-Carter, D. (2004). *Quantitative psychological research: A student's handbook*. New York: Psychology Press.

DeWall, C. N., & Maner, J. K. (2008). High status men (but not women) capture the eye of the beholder. *Evolutionary Psychology, 6,* 328–341.

Eagly, A. H., Makhijani, M .G., & Klonsky, B.G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111,* 3-22.

Eagly, A.H., & Steffens, V. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46,* 735–754.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology, 69*, 275-298.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22,* 133-187.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160.

Fiske, S. T., & Taylor, S, E. (2010). *Social cognition: From brains to culture.* New York: McGraw Hill.

Fried, R., Schettlinger, K., & Borowski, M. (2014). robfilter: Robust Time Series Filters. [Computer software]. Retrieved from https://CRAN.R-project.org/package=robfilter

Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which eye tracker is right for your research? Performance evaluation of several cost variant eye trackers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60,* 1240-124).

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology, 85,* 453-466.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin, 43*, 300-312.

Giessner, S. R., & Schubert, T.W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes, 104,* 30-44.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., Cohen, J. D. (2010) Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience, 10,* 252-69.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. *Journal of Personality and Social Psychology, 97*, 17–41.

Guinote, A. (2017). How power affects people: Activating, wanting and goal seeking. *Annual Review of Psychology, 68*, 353–381.

Hanel, P., Maio, G., & Manstead, A.S.R. (2019). A new way to look at the data: Similarities between groups of people are large and important. *Journal of Personality and Social Psychology, 116*, 541-562.

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R.E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition and Emotion, 14*, 609-642.

Hamilton, H. W., & Deese, J. (1971). Does linguistic marking have a psychological correlate? *Journal of Verbal Learning and Verbal Behavior, 10*, 707–714.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446.

Jamrozik, A., McQuire, *M*., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin and Review, 23,* 1080-1089.

Jarvis, B.G. (2012). DirectRT (Version v2012) [Computer Software]. New York, NY: Empirisoft Corporation.

Jost, J. T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable in social and political psychology. *Current Directions in Psychological Science*, *28*, 1-10.

Judd, C. M., Westfall, J., & Kenny D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54-69.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., . . . Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist,* 1-10.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing, 12*, 13–21.

Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science, 8*, 875-881.

Lakoff, G., & Johnson, *M*. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic Books.

Lebois, A.M, Wilson-Mendenhall, C.D., & Barsalou, L.W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science, 39,* 1764-1801.

Lewin, K. (1941). Regression, retrogression and development. *University of Iowa Studies of Child Welfare*, 18, 1–43.

Lin, H., Saunders, B., Hutcherson, C. A., & Inzlicht, M. (2018). Midfrontal theta and pupil dilation parametrically track subjective conflict (but also surprise) during intertemporal choice. *NeuroImage, 172*, 838-852.

Maass, A., Suitner, C., Favaretto, X., & Cignacchi, M. (2009). Groups in space: Stereotypes and the spatial agency bias. *Journal of Experimental Social Psychology, 45*, 496–504.

Mattan, B. D., Kubota, J. T., & Cloutier, J. (2017). How social status shapes person perception and evaluation: A social neuroscience perspective. *Perspective on Psychological Science, 12,* 468-507.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods, 50*, 94-106.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*, 1469-1481.

Meier, B.P., & Robinson, M.D. (2004). Why the sunny side is up. *Psychological Science, 15*, 243-247.

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C., K., & Schjeldahl, K (2007). *Journal of Personality and Social Psychology, 93*, 699-710.

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with white and black men. *Personality and Social Psychology Bulletin, 28,* 939–952.

Montepare, J.M. (1995). The impact of variations in height on young children's impressions of men and women. *Journal of Nonverbal Behavior*, *19*, 31-47.

Moor, G. B., Crone, E. A., & van der Molen, M. W. (2010). The heart brake of social rejection: Heart rate deceleration in response to unexpected peer rejection. *Psychological Science, 21*, 1326–1333.

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology, 48*, 1532-1543.

Oliveira, F. T. P., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience, 19*, 1994–2004.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171-192.

Payne, B. K., Vuletich, H. A., & Lunberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*, 233-248.

Preuschoff, K., Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience, 5*, 115.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*, 416 – 442.

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences, 16,* 285–291.

Proulx, T., Sleegers, W. and Tritt, S. M. 2017. The expectancy bias: Expectancy-violating faces evoke earlier pupillary dilation than neutral or negative faces. *Journal of Experimental Social Psychology, 70*, 69-79.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2016). Retrieved from https://www.pstnet.com.

R Core Team (2016). R: A language and environment for statistical computing [Computer software]. Retrieved from https://www.R-project.org.

Robinson, M. D., Zabelina, D. L., Ode, S., & Moeller, S. K. (2008). The vertical nature of dominance-submission: Individual differences in vertical attention. *Journal of Research in Personality, 42*, 933-948.

Roets, A., van Hiel, A., & Dhont, K. (2012). Is sexism a gender issue? A motivated social cognition perspective on men's and women's sexist attitudes toward own and other gender. *European Journal of Personality, 26,* 350-359.

Rondeel, E. W. M., van Steenbergen, H., Holland, R. W., & van Knippenberg, A. (2015). A closer look at cognitive control: Differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience, 9,* 494.

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*, 331-363.

Rudman, L.A., & Kilianski, S.E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin, 26,* 1315-1328.

Rudman, L.A., Moss-Racusin, C.A., Phelan, J.E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology, 48,* 165-179.

Rudman, L.A., & Phelan, J.E. (2010). The effects of priming gender roles on women's implicit gender beliefs and career aspirations. *Social Psychology, 41,* 192-202.

Schubert, T. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89,* 1-21.

Sherman, G. D., & Clore, G. L. (2009). The color of sin: White and black are perceptual symbols of moral purity and pollution. *Psychological Science, 20*, 1019-1025

Sleegers, W.W.A., Proulx, T., & van Beest, I. (2015). Extremism reduces conflict arousal and increases values affirmation in response to meaning violations. *Biological Psychology, 108,* 126–131.

Smith, L. & Gasser, M. (2006). The development of embodied cognition: Six lessons from babies. *Artificial Life, 11,* 13-29.

Smallwood, J., Brown, K.S., Tipper, C., Giesbrecht, B., Franklin, M.S., Mrazek, M.D., et al. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE 6,* e18298.

Spatola, N., Santiago, J., Beffara, B., Mermillod, M., Ferrand, L., & Ouellet, M. (2018). When the sad past is left: The mental metaphors between time, valence, and space. *Frontiers in Psychology, 9,* 1019.

Sriram, N., & Greenwald, A.G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*, 283-294.

Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General, 141*, 715-727.

Suitner, C., Maass, A., & Ronconi, L. (2017). From spatial to social asymmetry: Spontaneous and conditioned associations of gender and space. *Psychology of Women Quarterly, 41*, 46-64.

Taylor, J. E., Lam, T., Chasteen, A., Pratt, J. (2015). Bow your head in shame, or, hold your head up with pride: Semantic processing of self-esteem concepts orients attention vertically. *PLoS ONE, 10,* e0137704.

Tobii Group. (2015, April 27). Retrieved May 14, 2018, from https://www.tobii.com.

van der Wel, P. & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review,* 1-11.

van Steenbergen, H., & Band, G. P. H. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience, 7,* 215.

von Hecker, U., Klauer, C.K. & Sankaran, S. (2013). Embodiment of social status: Verticality effects in multi-level rank-orders. *Social Cognition, 31,* 374-389.

von Hecker, U., Klauer, K.C., Wolf, L., & Fazilat-Pour, *M*. (2016). Spatial processes in linear ordering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1003-1033.

Westfall, J. (2016). PANGEA: Power ANalysis for GEneral Anova designs. University of Texas at Austin: Unpublished manuscript.

Wood, W., & Eagly, A.H. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychological Bulletin, 128*, 699-727.

Zanolie, K., van Dantzig, S., Boot, I., Wijnenc, J., Schubert, T., Giessner, S., Pecher, P. (2012). Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical dimension. *Brain and Cognition, 78,* 50-58.

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. *Neuron, 58*, 273-283.

Appendix A

Procedure of modelling effects

In order to establish the random structure of the final linear mixed model for each study, we first estimated a minimal model, in which the intercepts varied across *participants.* Subsequently, we estimated three other models that had a similar structure to the minimal model, except that in each model, we also introduced one random slope to test whether random effects improved the fit of our data in the final model. That is, in Model 1, we introduced *gender choice* (men and women as powerful and powerless) as a random slope, in Model 2, we included *trial type* (men-top or women-top). After estimating the models, we compared the fit of each model that included a random slope with the minimal model by using Chi-square difference statistic $\Delta\chi 2$. If a model with a random slope provided a better fit for our data (i.e., the loglik ratio was significantly smaller than the ratio of the minimal model), the given random slope was then kept in the final model. The comparisons between minimal and random slope models as well as the structure of final models for each study are presented in tables below.

Study 1.

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	10	42157	42217	-21069	42137			
Model 1 Gender choice	12	42156	42228	-21066	42132	5.54	2	.06
Model 2 Trial type	12	42147	42219	-21062	42123	14.24	2	.001*

The final model included *trial type* as a random slope.

Study 2.

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	10	40866	40925	-20423	40846			
Model 1 Gender choice	12	40868	40939	-20422	40844	1.54	2	.46
Model 2 Trial type	12	40869	40941	-20423	40845	.50	2	.78

The final model did not include any random slopes.

Study 3.

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	10	27885	27941	-13933	27865			
Model 1 Gender choice	12	27883	27950	-13930	27859	6.01	2	.05*
Model 2 Trial type	12	28048	27889	-27956	27865	.01	2	.99

The final model included *gender choice* as a random slope.

Study 4.

Model	df	AIC	BIC	loglik	deviance	Δχ2	Δdf	*p*
Minimal Model	18	60161	60275	-30062	60125			
Model 1 Gender choice	20	60153	60280	-30057	60113	11.65	2	.01*
Model 2 Trial type	20	60164	60290	-30062	60124	1.0	2	.61
Model 3 Prime	20	60162	63000	-30061	60122	2.60	2	.27

The final model included *gender choice* as random slope.

Study 5.

Model	df	AIC	BIC	loglik	deviance	Δχ2	Δdf	*p*
Minimal Model	10	33065	33122	-16522	33045			
Model 1 Gender choice	12	33023	33092	-16500	32999	45.80	2	.001*
Model 2 Trial type	12	33069	33138	-16522	33045	.16	2	.93

The final model included *gender choice* as a random slope.

Study 6.

Model	df	AIC	BIC	loglik	deviance	Δχ2	Δdf	*p*
Minimal Model	10	32471	32529	-16226	32451			
Model 1 Gender choice	12	32471	32539	-16223	32447	4.64	2	.99
Model 2 Trial type	12	32473	32542	-16225	32449	3.01	2	.37

The final model did not include any random slopes.

Studies 1-6 – integrative analysis

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	50	238639	239025	-119270	238539			
Model 1 Gender choice	52	238599	238999	-119247	238495	44.90	2	.001*
Model 2 Trial type	52	239042	239042	-119268	238537	2.70	2	.26

The final model included *gender choice* as a random slope.

Study 7.

(a) Schubert block (overall analysis)

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	6	184	214	-85	172			
Model 1 Position	8	184	225	-84	168	3.16	2	.21
Model 2 Response period	8	155	155	-49	98	73.0	2	.001*

The final model included *response period* as a random slope.

(b) Gender block (overall analysis)

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	6	115	166	-48	95			
Model 1 Gender choice	8	105	166	-40	81	14.41	2	.001*
Model 2 Trial type	8	118	179	-47	94	1.32	2	.52
Model 3 Response period	8	35	97	-6	11	84.11	2	.001*

Appendix B

Method of individual studies (1 – 6)

Studies 1 and 2

Materials

Spatial task. We selected ten popular British (Study 1) and Polish (Study 2) names (five female and five male).[23] Each name was randomly paired with a name of the opposite gender (e.g., Oliver – Emily). The matched pairs were then presented on the computer screen in white letters on a black screen (font size 15 – 21 points).

Design

Spatial task. The same design was used in both studies. Task condition (find powerful versus powerless) was manipulated between-participants while the trial type (men-top versus women-top) was manipulated within-participants. There were four blocks of trials. Each block included a presentation of five pairs shown twice (10 trials in total). In half of the trials, a male name (e.g., Oliver) was displayed at the top of the screen with a female name (e.g., Emily) at the bottom. Both names were centered and there was a 23cm vertical distance between the names. In the other trials, this display was reversed. There were 40 trials in total. The order of trials within each block was randomized.

Social Status - Gender IAT. We adapted the IAT of implicit attitudes towards high or low status men and women from research by Rudman and Kilianski (2000). Within a single block participants were asked to categorize items that belonged to either high status/men or low status/women, or vice versa in another block. The order of the blocks was counterbalanced across participants.

[23] See the most popular names: Popular British baby names: Year by year. (n.d.). Retrieved February 9, 2015, from http://www.babycentre.- co.uk/popular- baby-names (Study 1); Academy of childbirth. (n.d.). Retrieved May 2, 2016, from: http://akademiaporodu.pl/top-news/najpopularniejsze-imiona-w-2015-2016-mapy-ranking (Study 2).

Procedure

The studies were presented using DirectRT (Jarvis, 2012). Participants sat approximately 70cm away from the computer screen. After reading the instructions, we verbally explained the concept of power to participants by adapting Galinsky et al. (2003) definition of social power. We explained the definition verbally to ensure that participants understood the definition of social power and to give them an opportunity to ask questions about it. Subsequently, participants completed the spatial task by pressing A or L on the computer keyboard to indicate the powerful/powerless person on the screen. In Study 2, the procedure was the same, but all instructions and stimuli were translated into Polish. After the spatial task, all participants completed the IAT.[24] Each experiment lasted approximately 15 minutes.

Study 3

Method

Materials

Spatial Task. The selected professions (derived from a pilot study) were assigned to male and female names (adapted from Study 1; e.g., Oliver-Professor, Sophie-Professor). Each pair (profession and name) was assigned to a corresponding pair including a name of member of the opposite gender. We created all combinations of professions and genders (e.g., male scientist - female professor; female professor - male scientist) within four sets of pairings. Participants were randomly assigned to receive one set of combinations.

Design

Spatial Task. We used the same mixed design as in Study 1 and 2 and presented participants with the same spatial task, except that in half the trials, presented in a random order, a male name and a profession were displayed at the top of the screen (e.g., Oliver-Professor) while a female name and a profession (e.g., Emily-Architect) at the bottom, whilst

[24] Within all the reported studies, participants signed consent forms at the beginning of experiments and were debriefed at the end.

for the remaining half of the trials, this spatial orientation was reversed. To facilitate responses, we asked participants to use arrows (up and down). Using these arrows provides a congruent mapping of mental simulations onto motor responses (see Schubert, 2005).

Social Status - Gender IAT. We used the same IAT of attitudes towards high status men and women, as in Study 1, but we reduced the number of blocks from five to two (see Sriram & Greenwald, 2009). Participants were asked to keep in mind two target categories within two blocks (men and high status in one block or women and high status in another) and respond as fast as possible by pressing a designated key when they saw an item that belonged to the target category on the screen. When they saw distractors (women and low status or men and low status) they were supposed to press another key. The order of blocks was counterbalanced across participants.

Procedure

First, participants completed the spatial task and then the IAT. Both tasks were programmed as in Studies 1 and 2. The experiment took approximately 15 minutes.

Study 4

Method

Materials

Spatial Task. We displayed the same gender pairs as in Study 1, but before each trial participants were presented with a social status item (gender-neutral profession adapted from Study 3, i.e.: scientist, architect, doctor, professor, dentist) or neutral word (i.e., vegetables: carrot, potato, lettuce, broccoli, cabbage). The prime was manipulated within-subjects. After making decisions about powerful/powerless person in each trial, participants were asked to report whether the initially presented prime word belonged to the social status (by pressing the arrow pointing left) or vegetable category (by pressing the arrow pointing right).[25] In this way, the prime word was activated in participants' minds during their decision about the powerful/powerless person.

Social Status - Gender IAT. We adapted the IAT from Study 1.

Design

Spatial Task. We used the same mixed design as in Study 1 and 2, and presented participants with a similar spatial task, except that before half of the trials (men-top; women-bottom) they were primed with a social status item, and before the other half (men-top; women-bottom) they were shown a neutral word. The same was done for the trials where women were presented at the top and men at the bottom, so there were 20 trials in total within each block. The trials were randomised within each block within-participants; there were 4 blocks in total.

Social Status - Gender IAT. We used the same IAT of gender authorities as in Study 1.

Procedure

First, participants completed the spatial task and then the IAT. The tasks were programmed using DirectRT (Jarvis, 2012). The experiment took 15 minutes.

Studies 5 and 6

Method

Materials and design

Spatial task. We used the same stimuli as in Study 1. However, participants responded with the arrow up to indicate the powerful/powerless person at the top or the arrow pointing down to indicate the target at the bottom in Study 5. Study 6 was a replication of Study 1, so participants responded with horizontally arranged keys on the keyboard.

[25] The responses were reversed for the other half of our participants: they pressed the arrow pointing right for social status items and the arrow pointing left for vegetables (participants were randomly assigned to one of the key arrangements).

Rationality/emotionality and gender IAT. We asked participants to quickly and accurately sort items that belonged to the category of emotionality and women as well as rationality and men and vice versa.

Rationality/emotionality and verticality IAT. Participants categorized items that belonged to the category of up and rationality and categories down and emotionality, or up/emotionality and down/rationality (the items for emotionality/rationality categories were the same as in the rationality/emotionality and gender IAT).

Both IATs were designed in the same standard way as the Social Status - Gender IAT used in Study 1.

Procedure

First, participants completed the spatial task, then the rationality/emotionality and gender IAT, which was followed by the IAT measuring the rationality/emotionality and verticality associations. We presented the tasks in the same order to all participants.

Appendix C

Studies 1, 3, 4, 5, 6: Five matched pairs of names

Pair 1. Oliver; Emily

Pair 2. Jack; Lily

Pair 3. Thomas; Chloe

Pair 4. Jacob; Alice

Pair 5. James; Sophie

Study 2: Five matched pairs of Polish names

Pair 1. Adam; Julia

Pair 2. Jakub; Emilia

Pair 3. Szymon; Alicja

Pair 4. Dawid; Marta

Pair 5. Piotr; Anna

Appendix D

Study 3: Four sets of assigned professions and genders

Set 1.

Pair 1. Oliver Professor; Emily Architect

Pair 2. Jack Doctor; Lily Scientist

Pair 3. Thomas Architect; Chloe Dentist

Pair 4. Jacob Dentist; Alice Doctor

Pair 5. James Scientist; Sophie Professor

Set 2.

Pair 1. Thomas Architect; Sophie Professor

Pair 2. James Scientist; Alice Doctor

Pair 3. Jacob Dentist; Emily Architect

Pair 4. Jack Doctor; Chloe Dentist

Pair 5. Oliver Professor; Lily Scientist

Set 3.

Pair 1. Jacob Dentist; Lily Scientist

Pair 2. Thomas Architect; Alice Doctor

Pair 3. James Scientist; Emily Architect

Pair 4. Jack Doctor; Sophie Professor

Pair 5. Oliver Professor; Chloe Dentist

Set 4.

Pair 1. Jacob Dentist; Sophie Professor

Pair 2. Jack Doctor; Emily Architect

Pair 3. Oliver Professor; Alice Doctor

Pair 4. James Scientist; Chloe Dentist

Pair 5. Thomas Architect; Lily Scientist

Appendix E

Studies 1 and 4: IAT categories and items

Female names: Karen, Caitlin, Jasmine, Charlotte, Ruby, Mary, Lauren, Kate, Zoe, Lara, Ann, Victoria, Bethany, Daisy, Sarah.

Male names: Brian, Kevin, Paul, Benjamin, Freddie, Joseph, Jake, Edward, Robert, Lewis, Toby, Liam, Patrick, Tommy, Arthur.

High status: Boss, supervisor, expert, leader, executive, authority.

Low status: Secretary, helper, aide, clerk, subordinate, assistant.

Study 2: IAT categories and items

Female names: Hanna, Paulina, Barbara, Justyna, Magdalena, Helena, Weronika, Klaudia, Dominika, Zofia, Zuzanna, Katarzyna, Aleksandra, Joanna, Oliwia.

Male names: Maciej, Jan, Filip, Wiktor, Gabriel, Marek, Konrad, Karol, Tomasz, Bartosz, Wojciech, Krzysztof, Patryk, Hubert, Adrian.

High status: Zwierzchnictwo, kierownictwo, nadzór, przywództwo, autorytet, dyrekcja.

Low status: Asysta, podporządkowanie, zależność, podrzędność, podwładność, służba.

Studies 5 and 6: IAT categories and items

Rationality: Intelligent, logic, reason, thinking.

Emotionality: Feeling, mood, sentiment, sympathy.

Up: Above, top, over, upper.

Down: Below, under, bottom, lower.

Female names: Karen, Caitlin, Jasmine, Charlotte, Ruby, Mary, Lauren, Kate, Zoe, Lara, Ann, Victoria, Bethany, Daisy, Sarah.

Male names: Brian, Kevin, Paul, Benjamin, Freddie, Joseph, Jake, Edward, Robert, Lewis, Toby, Liam, Patrick, Tommy, Arthur

Appendix F

Table 3
Summary of the aims and findings of each individual behavioral study (Studies 1 - 6).

Study	Specific aims	Findings
Study 1	To explore whether people spatially represent men as powerful at the top and women as powerless at the bottom when making judgements about gendered names.	1) Powerful-men-top (969 ms; 95% CI[848, 1090]) vs. powerful-men-bottom (1040 ms; 95% CI[904, 1176]), $p < .002$. 2) Powerful-men-top vs. powerful-women-top (1086 ms; 95% CI[948, 1225]), $p < .01$. 3) Powerless-women-bottom (998 ms; 95% CI[879, 1116]) vs. powerless-women-top (998 ms; 95% CI[866, 1131]), $p = 1$. 4) Powerless-women-bottom vs. powerless-men-bottom (1017 ms; 95% CI[883, 1151]), $p = .880$.
Study 2	To explore the same processes as in Study 1, however, by using a sample of Polish as opposed to British students.	1) Powerful-men-top (1143 ms; 95% CI[1031, 1255]) vs. powerful-men-bottom (1194 ms; 95% CI[1081, 1307]), $p = .130$. 2) Powerful-men-top vs. powerful-women-top (1218 ms; 95% CI[1106, 1330]), $p < .004$. 3) Powerless-women-bottom (1208 ms; 95% CI[1087.021 1330.375]) vs. powerless-women-top (1192 ms; 95% CI[1080, 1305]), $p = .960$. 4) Powerless-women-bottom vs. powerless-men-bottom (1224 ms; 95% CI[1108, 1339]), $p = .970$.
Study 3	To determine whether the spatial representation of gender would be more pronounced when people think about gender in the context of social-status cues.	1) Powerful-men-top (1175 ms; 95% CI[1055, 1256]), vs. powerful-men-bottom (1175 ms; 95% CI[1055, 1256]), $p < .002$. 2) Powerful-men-top vs. powerful-women-top (1108 ms; 95% CI[996, 1221]), $p = .710$. 3) Powerless-women-bottom (1258 ms; 95% CI[1143, 1373]) vs. powerless-women-top (1289 ms; 95% CI[1175, 1404]), $p = .660$. 4) Powerless-women-bottom vs. powerless-men-bottom (1288 ms; 95% CI[1186, 1390]), $p = .730$.
Study 4	To experimentally determine whether social status cues affect the strength of spatial simulations by directly priming participants with social-status items on some of the spatial task trials.	1) Powerful-men-top (1010 ms; 95% CI[895, 1126]) vs. powerful-men-bottom (1054 ms; 95% CI[938, 1170]), $p < .088$. 2) Powerful-men-top vs. powerful-women-top (1079 ms; 95% CI[957, 1201]), $p < .056$. 3) Powerless-women-bottom (1048 ms; 95% CI[932, 1165]) vs. powerless-women-top (1033 ms; 95% CI[916, 1150]), $p = .856$. 4) Powerless-women-bottom vs. powerless-men-bottom (1047 ms; 95% CI[934, 1160]), $p = .989$.
Study 5	To replicate the results of Study 1 when participants responded with vertically arranged keys on the keyboard.	1) Powerful-men-top (979 ms; 95% CI[896, 1062]) vs. powerful-men-bottom (996 ms; 95% CI[912, 1079]), $p = .850$. 2) Powerful-men-top vs. powerful-women-top (1040 ms; 95% CI[931, 1150]), $p = .450$. 3) Powerless-women-bottom (1093 ms; 95% CI[992, 1195]) vs. powerless-women-top (1112 ms; 95% CI[1010, 1214]), $p = .830$. 4) Powerless-women-bottom vs. powerless-men-bottom (1125 ms; 95% CI[1033, 1216]), $p = .860$.
Study 6	To replicate the results of Study 1 when participants responded with horizontally arranged keys on the keyboard.	1) Powerful-men-top (1010 ms; 95% CI[865, 1155]) vs. powerful-men-bottom (1012 ms; 95% CI[866, 1157]), $p = .999$. 2) Powerful-men-top vs. powerful-women-top (1002 ms; 95% CI[855, 1149]), $p = .999$. 3) Powerless-women-bottom (1043 ms; 95% CI[898, 1188]) vs. powerless-women-top (1037 ms; 95% CI[891, 1183]), $p = 1$.

		4) Powerless-women-bottom vs. powerless-men-bottom (1093 ms; 95% CI[946, 1240]), p = .310.